QUT

THE DEVELOPMENT OF RAPID GENOTYPING METHODS FOR METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS*

A thesis in
Molecular Microbiology
by

Alexander John Stephens

BAppSc (Hons)

CRC for Diagnostics

School of Life Sciences, Institute of Health and Biomedical Innovation

Queensland University of Technology

Brisbane, Australia

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy of the Queensland University of Technology

2008

ABSTRACT

Methicillin-resistant *Staphylococcus aureus* (MRSA) is an important human pathogen that is endemic in hospitals all over the world. It has more recently emerged as a serious threat to the general public in the form of community-acquired MRSA. MRSA has been implicated in a wide variety of diseases, ranging from skin infections and food poisoning to more severe and potentially fatal conditions, including; endocarditis, septicaemia and necrotising pneumonia. Treatment of MRSA disease is complicated and can be unsuccessful due to the bacterium's remarkable ability to develop antibiotic resistance.

The considerable economic and public health burden imposed by MRSA has fuelled attempts by researchers to understand the evolution of virulent and antibiotic resistant strains and thereby improve epidemiological management strategies. Central to MRSA transmission management strategies is the implementation of active surveillance programs, via which unique genetic fingerprints, or genotypes, of each strain can be identified. Despite numerous advances in MRSA genotyping methodology, there remains a need for a rapid, reproducible, cost-effective method that is capable of producing a high level of genotype discrimination, whilst being suitable for high throughput use. Consequently, the fundamental aim of this thesis was to develop a novel MRSA genotyping strategy incorporating these benefits.

This thesis explored the possibility that the development of more efficient genotyping strategies could be achieved through careful identification, and then simple interrogation, of multiple, unlinked DNA loci that exhibit progressively increasing mutation rates. The baseline component of the MRSA genotyping strategy described in this thesis is the allele-specific real-time PCR interrogation of slowly evolving core single nucleotide polymorphisms (SNPs). The genotyping SNP set was identified previously from the Multi-locus sequence typing (MLST) sequence database using an in-house software package named Minimum SNPs. As discussed in Chapter Three, the genotyping utility of the SNP set was validated on 107 diverse Australian MRSA isolates, which were largely clustered into groups of related strains as defined by MLST. To increase the resolution of the SNP genotyping method, a

selection of binary virulence genes and antimicrobial resistance plasmids were tested that were successful at sub typing the SNP groups.

A comprehensive MRSA genotyping strategy requires characterisation of the clonal background as well as interrogation of the hypervariable Staphylococcal Cassette Chromosome *mec* (SCC*mec*) that carries the β-lactam resistance gene, *mec*A. SCC*mec* genotyping defines the MRSA lineages; however, current SCC*mec* genotyping methods have struggled to handle the increasing number of SCC*mec* elements resulting from a recent explosion of comparative genomic analyses. Chapter Four of this thesis collates the known SCC*mec* binary marker diversity and demonstrates the ability of Minimum SNPs to identify systematically a minimal set of binary markers capable of generating maximum genotyping resolution. A number of binary targets were identified that indeed permit high resolution genotyping of the SCC*mec* element. Furthermore, the SCC*mec* genotyping targets are amenable for combinatorial use with the MLST genotyping SNPs and therefore are suitable as the second component of the MRSA genotyping strategy.

To increase genotyping resolution of the slowly evolving MLST SNPs and the SCC*mec* binary markers, the analysis of a hypervariable repeat region was required. Sequence analysis of the Staphylococcal protein A (*spa*) repeat region has been conducted frequently with great success. Chapter Five describes the characterisation of the tandem repeats in the *spa* gene using real-time PCR and high resolution melting (HRM) analysis. Since the melting rate and precise point of dissociation of double stranded DNA is dependent on the size and sequence of the PCR amplicon, the HRM method was used successfully to identify 20 of 22 *spa* sequence types, without the need for DNA sequencing.

The accumulation of comparative genomic information has allowed the systematic identification of key MRSA genomic polymorphisms to genotype MRSA efficiently. If implemented in its entirety, the strategy described in this thesis would produce efficient and deep-rooted genotypes. For example, an unknown MRSA isolate would be positioned within the MLST defined population structure, categorised based on its SCC*mec* lineage, then subtyped based on the polymorphic *spa* repeat region. Overall, by combining the genotyping methods described here, an integrated and novel

MRSA genotyping strategy results that is efficacious for both long and short term investigations. Furthermore, an additional benefit is that each component can be performed easily and cost-effectively on a standard real-time PCR platform.

LIST OF KEY WORDS

Methicillin-resistant *Staphylococcus aureus*, MRSA, single nucleotide polymorphism, SNP, Multi-locus sequence typing, MLST, genotyping, high resolution melt, HRM, Minimum SNPs, real-time PCR, Staphylococcal cassette chromosome *mec*, SCC*mec*, Staphylococcal protein A, *spa*, comparative genomics, bioinformatics, binary markers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

agr   Accessory genome regulator

AGRF   Australian Genome Research Facility

arcC   Carbamate kinase gene

aroE   Shikimate dehydrogenase gene

AFLP   Amplified fragment length polymorphisms

AS   Allele-specific

BHI   Brain heart infusion

bp   Base pair(s)

CA-MRSA Community acquired Methicillin-resistant *Staphylococcus aureus*

CC   Clonal complex

CGH   Comparative genomic hybridisation

CoNS   Coagulase-negative Staphylococci

CT   Cycle time

DLV   Double locus variant(s)

DNA   Deoxyribonucleic acid

ECM   Extracellular matrix

glp   Glycerol kinase gene

gmk   Guanylate kinase gene

HRM   High resolution melt

HVR   Hyper variable region

Kb   Kilo base(s)

MALDI-TOF MS Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry

Mbps   Mega base pairs

MSCRAMM Microbial surface components recognising adhesive matrix molecules

MGE   Mobile genetic element(s)

mL   Milli Litre(s)

MLST   Multi-locus sequence typing

MLVA   Multiple locus VNTR Analysis

mm   Millimetre(s)

MREJ   *mec* right extremity junction

MRSA	Methicillin-resistant *Staphylococcus aureus*
MSSA	Methicillin-sessitive *Staphylococcus aureus*
ng	Nano grams(s)
ORF	Open reading frame
SaPI	*Staphylococcus aureus* pathogenicity island
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PHRANA	Progressive hierarchical resolving assays using nucleic acids
pmol	Pico mole(s)
pta	Phosphate acetyltransferase gene
pvl	Panton Valentine leukocidin
SCC*mec*	Staphylococcal Cassette Chromosome *mec*
SLV	Single locus variant(s)
SNP	Single nucleotide polymorphism
spa	Staphylococcal protein A
SSRs	Short Sequence Repeats
ST	Sequence type
T_m	Melting temperature
tpi	Triosephosphate isomerise gene
VISA	Vancomycin intermediate *Staphylococcus aureus*
VNTR	Variable number of tandem repeats
VRSA	Vancomycin resistant *Staphylococcus aureus*
yqil	Acetyl coenzyme A acetyltransferase gene
μL	Micro litre(s)

LIST OF PUBLICATIONS

REFEREED JOURNAL PUBLICATIONS

Methicillin-resistant *Staphylococcus aureus* genotyping using a small set of polymorphisms.

Stephens AJ, Huygens F, Inman-Bamber J, Price EP, Nimmo GR, Schooneveldt J, Munckhof W, Giffard PM.

J Med Microbiol. 2006 Jan; 55 (Pt 1):43-51.

Systematic derivation of marker sets for Staphylococcal Cassette Chromosome *mec* typing.

Stephens AJ, Huygens F and Giffard PM

Antimicrob Agents Chemother. 2007 Aug;51(8):2954-64.

High-Resolution Melting analysis of the *spa* repeat region of *Staphylococcus aureus*.

Stephens AJ, Giffard PM, Inman-Bamber J and Huygens F

Clin Chem. 2008 Feb;54(2):432-6.

Use of a single-nucleotide polymorphism genotyping system to demonstrate the unique epidemiology of methicillin-resistant *Staphylococcus aureus* in remote aboriginal communities.

McDonald M, Dougall A, Holt D, Huygens F, Oppedisano F, Giffard PM, Inman-Bamber J, Stephens AJ, Towers R, Carapetis JR, Currie BJ.

J Clin Microbiol. 2006 Oct;44(10):3720-7.

Carriage of methicillin-resistant *Staphylococcus aureus* in a Queensland Indigenous community.

Vlack S, Cox L, Peleg AY, Canuto C, Stewart C, Conlon A, Stephens A, Giffard P, Huygens F, Mollinger A, Vohra R, McCarthy JS.

Med J Aust. 2006 Jun 5;184(11):556-9.

Life-threatening community-acquired methicillin-resistant *Staphylococcus aureus* infection in Australia.

Peleg AY, Munckhof WJ, Kleinschmidt SL, Stephens AJ, Huygens F.

Eur J Clin Microbiol Infect Dis. 2005 Jun;24(6):384-7.

CONFERENCE PRESENTATIONS

Binary marker typing increases resolution of methicillin-resistant *Staphylococcus aureus* MLST clonal groups defined by SNP typing

Alex J Stephens, Flavia Huygens, Graeme R Nimmo and Philip M Giffard

Poster, International Symposium on Staphylococci and Staphylococcal Infections (ISSSI), Charleston, 2004

SCC*mec* genotyping using Minimum SNPs

Alex J Stephens, Flavia Huygens and Philip M Giffard

Poster, Australian Society of Microbiology annual conference, Gold Coast 2006

PATENT

Giffard, P. M., F. Huygens, E. P. Price, A. J Stephens and J. Inman-Bamber

A Diagnostic method (Provisional). 2006905879. Filed 23-10-06

STATEMENT OF ORIGINAL AUTHORSHIP

The work presented in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Signed: …………………………………

Alexander John Stephens
BAppSci (Hons)

Date: ……………………

ACKNOWLEDGEMENTS

Foremost, I would like to thank my primary supervisor Assoc. Prof. Phil Giffard and my associate supervisor Dr. Flavia Huygens, who throughout the duration of my project remained dedicated, attentive, interested and patient. They provided an excellent environment to explore the wonderful world of bacteria and were a pleasure to be around. Thank you for all the time and effort you have put into my scientific development, for which I am truly grateful.

Secondly, I would like to acknowledge my wife Renee, who has been amazingly supportive and understanding during our PhD journey. Thank you so much for pushing me when I needed a shove and pulling me back when I needed a break. You lightened the burden right from the start and carried it until the end. I can only try to repay my debt to you.

Thirdly, I would like to acknowledge my colleagues in the CRC for Diagnostics for all the wonderful times spent both inside and outside the laboratory. Your support and friendship has made this period of my life unforgettable.

Lastly, I would like to thank my parents Bob and Lorna, who although don't really understand what I have been up to all these years, have nonetheless supported me in my endeavours. Thank you for teaching me the importance of education.

CHAPTER ONE

LITERATURE REVIEW

1.1 INTRODUCTION

Staphylococcus aureus is a major cause of health-care facility and community-acquired disease. It is now well established that the transmission of Staphylococcal disease can be reduced by stringent public health and infection control measures. Consequently, there is ample justification for monitoring the dissemination of these organisms at all scales, from within individual health care facilities to between continents. Tracking bacterial dissemination requires assigning isolates to subgroups within the species. This is known as typing, and if the typing is based on genetic analysis, it is usually known as genotyping. An ideal *S. aureus* genotyping method would be rapid, cost effective, and be easily portable among laboratories. It should also provide high informative power in relation to the population diversity and structure that is easily understood. This is necessary for reliable testing of hypotheses regarding epidemiological linkage. It also potentially facilitates drawing inferences from genotypes regarding virulence and resistance phenotypes. Developing such a genotyping method was the fundamental objective of this project.

1.2   *STAPHYLOCOCCUS AUREUS*

1.2.1   *General Characteristics*

Staphylococcus aureus is a prominent human pathogen that was first reported in 1881 by Alexander Ogston (Ogston, 1881). After microscopic analysis of purulent infections, Ogston discovered grape-like clusters of round, golden cells. Following this observation he replicated the original infections by injecting experimental subjects with the putative infectious organism. It is now established that *S. aureus* is a Gram-positive, facultative aerobe that produces catalase and coagulase and can tolerate reduced water activity (Madigan *et al.*, 2000b). Under optimal conditions, cell division occurs approximately every 20min with cell diameter ranging from 0.5-1.5μm. *S. aureus* is able to grow on a range of nutrient media and appears as smooth, convex colonies with diameters of ~1-3mm. The golden pigmentation, caused by carotenoids, is usually visible, and this is thought to protect cells against ultra-violet radiation (Allegra *et al.*, 1954).

1.3   *S. AUREUS* ANTIBIOTIC RESISTANCE

1.3.1   *Penicillin Resistance*

Penicillin was discovered serendipitously in 1928 and is lethal towards sensitive cells via specific deactivation of cell-wall-located penicillin binding protein (PBP) transpeptidases (Fleming, 1929). Inactivated PBPs are prevented from cross linking the peptidoglycan stands that in turn leads to a weakened cell wall and subsequent death by osmosis. The implementation of penicillin as a frontline clinical treatment for *S. aureus* infection did not occur until the early 1940s. Before this time, *S. aureus* infections could not be treated by means other than bandaging and basic wound care. By the 1950s, penicillin was a well established "wonder drug" and encouraged the wide notion in the public arena and medical profession that the battle against infectious diseases would be won within their generation. However, this did not happen. Isolates that acquired a penicillin hydrolysing enzyme (penicillinase) were favoured by natural selection and resistant strains arose (Spink and Ferris, 1945).

1.3.2   Methicillin Resistance

To combat the spread of penicillin-resistant *S. aureus* strains, the penicillin molecule was chemically modified for immunity against penicillinase activity. The resulting molecule is now known as methicillin or meticillin (Rolinson *et al.*, 1960). By demonstrating resistance to penicillinase and PBP binding affinity, methicillin was hailed as the definitive treatment against *S. aureus* infection. But in 1961, within 12 months of its clinical release, the first treatment failure occurred and methicillin-resistant *S. aureus* strains (MRSA) had evolved (Jevons, 1961). Resistance to methicillin and all β-lactam antibiotics is mediated by an exogenous substitute for the intrinsic PBPs, termed PBP2` or PBP2a, which is carried by a mobile genetic element termed the Staphylococcal cassette chromosome *mec* (SCC*mec*). PBP2a features subtle structural changes that reduce the binding affinity of methicillin, so avoiding inactivation and therefore maintaining a viable cell wall in otherwise lethal antibiotic conditions (Hartman and Tomasz, 1984; Hayes, 1981).

1.3.3   Vancomycin Resistance

The most effective treatment presently available against multi-resistant MRSA infections is the glycopeptide vancomycin; however, like penicillin and its derivatives, vancomycin resistance has also evolved. Isolates demonstrating low to moderate levels of resistance are termed vancomycin-intermediate *S. aureus* (VISA) and have been reported since 1996 from various locations, including Japan (Hiramatsu *et al.*, 1997a; Hiramatsu *et al.*, 1997b), United States (Smith *et al.*, 1999) and Europe (Ploy *et al.*, 1998). While the mechanism of resistance has not been fully elucidated, observations of abnormally thick cell walls indicate that increased cell wall synthesis may be important (Hanaki *et al.*, 1998; Hiramatsu *et al.*, 1997b). It has been hypothesised that the excess peptidoglycan ensnares the vancomycin molecule within the cell wall, therefore preventing access to the cytoplasmic target of the N-acetyl-muramic acid precursor (Cui *et al.*, 2000; Hiramatsu, 1998; Hiramatsu and Hanaki, 1998; Hiramatsu, 2001). The first high level vancomycin-resistant *S. aureus* (VRSA) isolate was reported from the United States in 2002 (Anonymous, 2002).

Unlike VISA, VRSA resistance is mediated by *vanA*, which is identical to the mechanism utilised by vancomycin-resistant *Enterococci* (VRE) (Weigel *et al.*, 2003; Woodford, 2001).

1.4 EPIDEMIOLOGY

An investigation into the epidemiology of *S. aureus* requires a sound understanding of the different aspects of this bacterium's transmission process and preferred environment. Below is a description of the factors involved as they relate to human disease.

1.4.1 Carriage

S. aureus is defined as a normal human commensal with the potential to cause opportunistic infections. The primary ecological niche of the species is the nasal inner wall opposite the nostril wing, otherwise known as the anterior nares or vestibulum nasi (Williams, 1963). The vestibulum nasi is composed of keratinised epithelial tissue, large apocrine sweat glands and is devoid of hair follicles or cilia. It has been demonstrated that keratinised epithelial tissue plays an important role in *S. aureus* adhesion (Bibel *et al.*, 1982). Other colonisation sites include the hands (Armstrong-Esther, 1976), gastrointestinal tract (Rimland and Roberson, 1986; Williams, 1963), vagina (Guinan *et al.*, 1982) and perineum (Ridley, 1959).

Carriage rates of *S. aureus* in humans are generally divided into three defined states: persistent carriage, intermittent carriage, and non-carriage (Armstrong-Esther, 1976; Vandenbergh *et al.*, 1999; Williams, 1963). It is generally accepted that in humans; 20% of the population are persistent carriers, 30% are intermittent carriers, and 50% are non-carriers (Eriksen *et al.*, 1995; Kluytmans *et al.*, 1997; Vandenbergh *et al.*, 1999, Wertheim, 2005 #1739). Surprisingly, MRSA carriage rates in the general public has been calculated to be as high as 7% (Cunningham *et al.*, 2007; Huang *et al.*, 2007; Hidron *et al.*, 2005), however, a meta-analysis of community prevalence

rates reported an average of approximately only 2% (Salgado *et al.*, 2003). The reported proportion of each carriage state varies depending on experimental variables such as the culture method, sites swabbed and population tested. Interestingly, persistent *S. aureus* carriers are often colonised with a single strain (Eriksen *et al.*, 1995; Vandenbergh *et al.*, 1999; Hu *et al.*, 1995), have higher bacterial loads, and are at increased risk of infection (Nouwen *et al.*, 2005; Nouwen *et al.*, 2004). Non-carriers produce antimicrobial nasal discharges with sufficiently strong activity to prevent colonisation (Cole *et al.*, 1999; Cole *et al.*, 2001).

1.4.2   Infection Risk Factors

Although *S. aureus* typically produce opportunistic infections, there are defined risk factors that increase this likelihood (Davis *et al.*, 2004). In general, the opportunistic infections occur when the host body's defences are weakened. Skin breakage and/or immunosuppression in combination with nasal carriage are the major risk factors (Kluytmans and Wertheim, 2005). The risk is amplified for residents of high risk environments such as hospitals and nursing homes (Storch *et al.*, 1987), particularly those associated with intensive care units (Donowitz *et al.*, 1982; Sista *et al.*, 2004), surgical procedures (Centofanti *et al.*, 2007; Desai *et al.*, 2003) and HIV infections (Smith *et al.*, 1994). Other conditions involving repeated or constant breakage of the skin barrier, such as haemodialysis treatment (Yu *et al.*, 1986) and intravenous drug abuse (Craven *et al.*, 1986) also increase the risk of developing serious *S. aureus* infections. Additional risk factors relating particularly to MRSA strains include prior antibiotic use, contact with MRSA infected or colonised persons and recent hospitalisation (Thompson *et al.*, 1982; Boyce, 1989). Hospitals with endemic MRSA focus their screening on high risk departments where MRSA is potentially most damaging (Huang *et al.*, 2006).

Low hygiene standards also contribute to higher levels of *S. aureus* infection. In particular, community-associated MRSA (CA-MRSA) carriage and therefore infection are associated with lower socioeconomic population groups (Cooke *et al.*, 2008; Ofner-Agostini *et al.*, 2006; Groom *et al.*, 2001; Charlebois *et al.*, 2002) and

children (Creech *et al.*, 2005). Children are prone to *S. aureus* infection as nose picking and minor scrapes and cuts are common. Unsurprisingly, children are able to alter their carriage states between the ages of 10 and 20 via behavioural change (Armstrong-Esther, 1976). Other at risk populations include prisoners (Baillargeon *et al.*, 2004), members of the male homosexual community (Lee *et al.*, 2005) and participants in body contact sports, specifically wrestling (Lindenmayer *et al.*, 1998) and football (Stacey *et al.*, 1998; Begier *et al.*, 2004). There is also evidence for transmission to humans from domestic pets (Boost *et al.*, 2007; Weese *et al.*, 2006) and food producing animals (Khanna *et al.*, 2008).

1.4.3 Disease

S. aureus is typically carried asymptomatically by humans and animals. It is responsible however, for a range of medical conditions that require treatment. The nature and extent of disease depends on the pathogenic characteristics of the infecting strain, host susceptibilities and the route of host entry. The most common ailments are skin and soft tissue infections that include abscesses, cellulitis, impetigo and postoperative surgical wounds (Cohen, 2007); with osteomyelitis also known to occur (Cunningham *et al.*, 1996). Strains carrying particular toxins can be associated with specific disease states, including staphylococcal scalded skin syndrome (SSSS) (Ladhani *et al.*, 1999), Toxic Shock Syndrome (TSS) (Novick and Subedi, 2007), food poisoning (Le Loir *et al.*, 2003), and necrotising pneumonia (Labandeira-Rey *et al.*, 2007). Septicaemia is a serious progression of *S. aureus* disease that occurs when the blood stream becomes infected. Between 1995 and 2002 20% of all nosocomial septicaemia cases in the United States were caused by *S. aureus* (Wisplinghoff *et al.*, 2004). In addition, *S. aureus* infections in the USA caused by MRSA {64% in 2003 (Klevens *et al.*, 2006)} have been linked statistically with higher mortality (Cosgrove *et al.*, 2003), longer hospital stays and elevated treatment costs compared with MSSA (Cosgrove *et al.*, 2005).

1.4.4   Prevention of Nosocomial Infection

As the great majority of antibiotic resistant and therefore serious *S. aureus* infections occur in healthcare facilities, transmission prevention is essential for reducing infection rates. Patient-to-patient transmission of MRSA occurs mainly via contaminated hands of health care professionals and their equipment, with aerial transmission also identified as a transmission route (Solberg, 2000). Risk of MRSA transmission can be minimised by following basic prevention protocols. These include; hand disinfection after patient contact, wearing gloves and gowns when treating MRSA affected patients, effective and rapid treatment of MRSA infections, and isolation of MRSA infected patients until three negative MRSA cultures are obtained (Verhoef *et al.*, 1999). Prudent use of broad spectrum antibiotics is also recommended for decreasing MRSA spread (Rosdahl and Knudsen, 1991). Screening patients at hospital admittance by evaluating the risk factors for MRSA colonisation is also an effective method for controlling transmission and infection (Zulian *et al.*, 1999).

Scandinavian countries and the Netherlands lead the world in nosocomial *S. aureus* infection control via strict implementation of effective treatment and containment measures (van Trijp *et al.*, 2007). MRSA infections as a percentage of all *S. aureus* hospital acquired infections are often under 5%, which has a dramatic effect on the safety of patients (Tiemersma *et al.*, 2004). In contrast, the majority of the world, and in particular developed countries including Japan and America, have widespread problems relating to infection prevention and control, with MRSA rates above 64% reported (Klevens *et al.*, 2006; Hijioka *et al.*, 2002). Surveys conducted by the Australian Group on Antimicrobial Resistance (AGAR) have calculated that 20% of *S. aureus* nosocomial isolates in Australia are MRSA (Nimmo *et al.*, 2003). High MRSA infection rates lead to complications with patient recovery that can ultimately lead to preventable deaths (Cosgrove *et al.*, 2003).

1.4.5   Community-acquired MRSA (CA-MRSA)

Community-acquired MSSA infections requiring medical attention can be easily treated with common antibiotics. Conversely, CA-MRSA infections are difficult to treat and pose a problem as general practitioners are often unaware of the transmission of MRSA within the general population. The notion that the epidemiology of MRSA was changing was exemplified by a report describing the MRSA associated deaths of four healthy children in 1997 (Anonymous, 1999). Since then, numerous CA-MRSA clones have been identified worldwide (Cohen, 2007), which has resulted in much research aimed at characterising CA-MRSA strains compared with their nosocomial relatives (Nimmo *et al.*, 2000; Fey *et al.*, 2003; Aires de Sousa *et al.*, 2005). In general, major differences noted thus far are that CA-MRSA are often non-multiresistant, carry small novel SCC*mec* elements (O'Brien *et al.*, 2005; Ma *et al.*, 2002) and carry the Panton-Valentine leukocidin (*pvl*) toxin (Diep *et al.*, 2004). The major threat posed by CA-MRSA is the replacement of MSSA as the human commensal, which is analogous to the historical propagation of penicillin-resistant strains.

1.5   *S. AUREUS* GENOME SEQUENCES

To gain a better understanding of the genetic attributes that allow *S. aureus* to display such diverse clinical phenotypes, the first fully sequenced and annotated genomes were published in 2001 (Kuroda *et al.*, 2001). The Japanese isolates, Mu50 and N315 were selected for whole genome comparison as they are closely related, but display considerable variation in regard to antibiotic resistance phenotypes. Since then, an additional 13 genomes have been sequenced and a further five are near completion (Table 1.1).

The completed *S. aureus* genomes are from diverse isolates with regard to antibiotic resistance and environmental source, however, they were primarily chosen due to their proven pathogenic capability. Of the 15 complete genomes, 10 are MRSA, three of which include vancomycin resistance and the remaining five are MSSA. The

genomes are between 2.74 and 2.90Mbps in size, and encompass between 2,560 and 2,892 open reading frames (ORFs) (both demonstrated experimentally and predicted). Of particular interest; MW2 (Baba *et al.*, 2002) and FPR3757 (Diep *et al.*, 2006) are representatives from major USA CA-MRSA clones, MRSA252 belongs to the dominant UK nosocomial clone (Johnson *et al.*, 2001), and COL is an early MRSA strain isolated in the 1960s (Gill *et al.*, 2005). In addition, the genomes of two isolates from a single patient were sequenced to demonstrate the stepwise genetic transition to vancomycin resistance (Mwangi *et al.*, 2007). The inclusion of a bovine isolate (RF122) provides an interesting comparison point as it is not likely to have been in contact with nosocomial environments (Herron-Olson *et al.*, 2007). This compilation of *S. aureus* genomes provides a sample of the variation within the species that is essential for understanding the genetic basis of complex physiological and pathological processes.

Table 1.1. The sequenced genomes of *S. aureus*.

Isolate	Type	Size (bp)	Status	GenBank accession no	Reference
RF122	Bovine MSSA	2,742,531	Complete	AJ938182	(Herron-Olson *et al.*, 2007)
COL	HA-MRSA	2,809,422	Complete	CP000046	(Gill *et al.*, 2005)
JH1	HA-MRSA	2,906,507	Complete	CP000736	(Mwangi *et al.*, 2007)
JH9	HA-VISA	2,906,700	Complete	CP000703	(Mwangi *et al.*, 2007)
MRSA252	HA-MRSA	2,902,619	Complete	BX571856	(Holden *et al.*, 2004)
MSSA476	CA-MSSA	2,799,802	Complete	BX571857	(Holden *et al.*, 2004)
MW2	CA-MRSA	2,826,402	Complete	BA000033	(Baba *et al.*, 2002)
Mu50	HA-VISA	2,878,529	Complete	BA000017	(Kuroda *et al.*, 2001)
N315	HA-MRSA	2,814,816	Complete	BA000018	(Kuroda *et al.*, 2001)
Mu3	HA-VISA	2,880,168	Complete	AP009324	(Neoh *et al.*, 2008)
NCTC8325	MSSA	2,821,361	Complete	CP000253	(Gillaspy, 2006)
FPR3757	CA-MRSA	2,872,769	Complete	CP000255	(Diep *et al.*, 2006)
Newman	MSSA	2,878,897	Complete	AP009351	(Baba *et al.*, 2008)
TCH1516	CA-MRSA	2,872,915	Complete	CP000730	(Highlander *et al.*, 2007)
TCH959	CA-MSSA	NA	Complete	NA	(Highlander *et al.*, 2007)
MSA553	MSSA	2,856,447	In progress	NA	Herron-Olson *et al.*, in prep.
D30	-	-	In progress	NA	(NCBI)
930918-3	-	-	In progress	NA	(NCBI)
0582	-	-	In progress	NA	(Sanger Institute)
EMRSA15	HA-MRSA	-	In progress	NA	(Sanger Institute)

NA = Not available.

1.6   GENOME STRUCTURE

S. aureus genomes display an overall conservation of sequence and structure; however this uniformity is disrupted with small pockets of heterogeneity (Lindsay and Holden, 2006). The conserved portion of the genome is generally comprised of genes responsible for common species function and is referred to as the "core" genome. The non essential sequence blocks not found uniformly in the species are deemed the "accessory" genome (Fig 1.1). Each genome component contains variation that accounts for the observed genotypic and phenotypic differences.

1.7   THE CORE GENOME COMPONENT

Both *in silico* and microarray analysis have determined that approximately 75% of the *S. aureus* genome is considered core sequence due to high conservation rates within the species (Lindsay and Holden, 2004; Fitzgerald *et al.*, 2001). Functional analysis of the ORFs from the N315 genome revealed that about half the proteins are involved in the vegetative life of the bacterium. Specifically, 18% of the genome ORFs are putatively involved in metabolic pathways, 9% are involved in the various transport systems, 6% in RNA synthesis, 4% in protein synthesis, and about 5% of ORFs encoded transcription regulators (Kuroda *et al.*, 2001). Approximately 75% of the ORFs involved in essential cell functions are most similar to *Bacillus subtilise* and *Bacillus halodurans*, a result suggesting that the *S. aureus* genome backbone is vertically derived from a common ancestor of the *Bacillus*/*Staphylococcus* group of bacteria (Kuroda *et al.*, 2001).
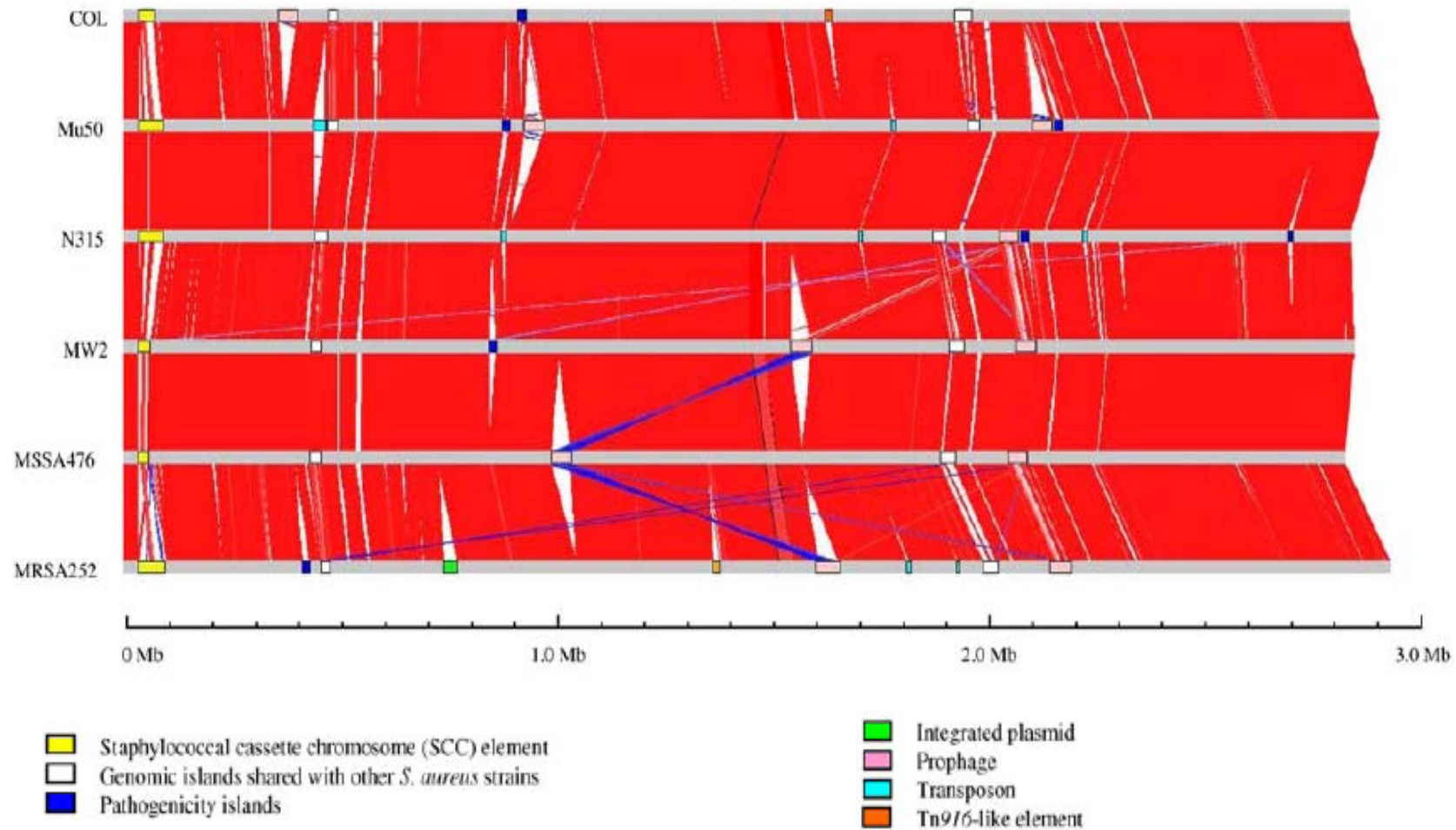
Fig 1.1. Comparison of multiple *S. aureus* genome sequences. The alignment of six *S. aureus* genomes sequences displays the largely syntenic structure (red) punctuated with blocks of sequence variation that comprise the accessory genome (colours listed in figure). The accessory genome is largely responsible for the various clinical phenotypes. Figure from Lindsay and co-worker (Lindsay and Holden, 2006).

The extent of core sequence conservation among *S. aureus* strains has also been determined. This was achieved by creating *in silico* "virtual chromosomes" that are devoid of polymorphic accessory sequences. The core genomes from related isolates (N315 and Mu50) displayed sequence homology of 99.7%, whereas MW2 had 94.8% and 94.7% with the unrelated N315 and Mu50, respectively (Baba *et al.*, 2002). This indicates that while highly conserved, the core genome features significant variation. This variation is accounted for by slowly accumulating single nucleotide polymorphisms (SNPs), hyper variable repeat regions and largely conserved yet non essential ORF that have been termed "core variable" (CV) genes (Lindsay *et al.*, 2006).

The CV genes are largely species-specific and include ORFs that encode cell surface binding proteins, variable numbers of tandem repeat loci and the accessory genome regulator (*agr*) (Lindsay and Holden, 2004). Physical interaction between *S. aureus* and its host is mediated by extracellular matrix (ECM) binding proteins that are covalently linked to the cell wall peptidoglycan (Navarre and Schneewind, 1994). These proteins are generally known as "microbial surface components recognising adhesive matrix molecules" (MSCRAMMs) (Patti and Hook, 1994), of which 20 are known to be produced by *S. aureus* (Mazmanian *et al.*, 2001). These proteins include, but are not limited to, the collagen adhesin protein (*cna*) (Patti *et al.*, 1992), the serine-aspartate dipeptide repeat family (*Sdr*C, D, E and *bbp*) (Josefsson *et al.*, 1998; Tung *et al.*, 2000), fibronectin binding proteins A and B (f*nbp*A, B) (Greene *et al.*, 1995), and clumping factor proteins A and B (*Clf*A, B) (McDevitt *et al.*, 1994; Ni Eidhin *et al.*, 1998). The MSCRAMMs contribute to core genome variation due to gene deletion, which is possible since they are dispensable.

Cell surface proteins also contribute to core genome variation as they feature variable number of tandem repeat (VNTR) domains. *S. aureus* proteins embedded in the cell wall share a similar protein structure consisting of three domains: the cell wall anchored domain, the linker domain and the ECM ligand-binding domain (Foster and Hook, 1998). Of particular interest due to their diversity in terms of content and length, is the linker domain that serves to expose the ligand-binding domain to the

ECM (Hartford *et al.*, 1997). This protein domain is comprised of tandem repeating amino acid units that are encoded by a corresponding DNA VNTR domain. These VNTR domains are subject to slip strand mispairing and consequently evolve at high mutation rates (van Belkum *et al.*, 1998a; Koreen *et al.*, 2004). Hypervariable regions appear to be the least stable aspect of the core genome and therefore play an important role in *S. aureus* genotype diversity (discussed in section 1.9.3).

The chimerical accessory gene regulator is a quorum-sensing system that regulates various colonisation and virulence factors and is also considered a CV component (Novick *et al.*, 1995). The *agr* consists of four components: *agr*A and *agr*C encode a signal transduction system (Lina *et al.*, 1998), whereas *agr*B and *agr*D are involved in production of an autoinducing peptide (AIP) for the quorum-sensing system (Ji *et al.*, 1995). The *agr* cluster is considered a CV locus as an internal region including the C terminus of *agr*B, all *agr*D and the N-terminal half of *agr*C displays sequence variation leading to production of four separate AIPs. This variation has led to evolution of four *agr* types that are defined by the ability of their corresponding AIPs to cross-activate or cross-inhibit *agr* quorum sensing (Ji *et al.*, 1997; McDowell *et al.*, 2001).

1.8 THE ACCESSORY GENOME

In contrast to the stable core, the accessory component accounts for most of the genetic variation within the species. Microarray experiments conducted by Fitzgerald *et al* estimated that ~22% of the *S. aureus* genome is variable in nature and is therefore not essential for survival (Fitzgerald *et al.*, 2001). These findings were corroborated and expanded by comparative analysis of multiple genome sequences (Baba *et al.*, 2002). These studies revealed that the accessory genome is composed of exogenous mobile genetic elements (MGE), whose horizontal transmission is fundamental to *S. aureus* evolution (Baba *et al.*, 2002; Fitzgerald *et al.*, 2001; Kuroda *et al.*, 2001). MGEs discovered to date include: bacteriophages, genomic islands, pathogenicity islands, transposons, plasmids and insertion sequences (Kuroda *et al.*,

2001). As *S. aureus* recombination and conjugation occurs rarely, the bulk of lateral MGE transfer is attributed to bacteriophage transduction.

Although MGEs are not compulsory for basic function, the diverse array of accessory genome constituents are strongly associated with mechanisms for adaptation to hostile environments, host colonization, and virulence processes (Fitzgerald *et al.*, 2001). Accordingly, identification and characterisation of MGEs has been of high importance in order to understand and control *S. aureus* infection. Furthermore, characterisation of the largely unlinked accessory components - in combination with the core genome - has provided a highly successful approach for generating high resolution genotypes, which is essential for differentiating closely related strains that share core genomes. The MGEs that contribute significantly to the diversity of *S. aureus* genotypes and phenotypes are described below.

1.8.1   *Transposable Elements*

Transposable elements are fragments of DNA that inherently encode the ability to transfer from one chromosomal site to another. Members of this MGE class characteristically harbour an integrase/transposase gene for mobility and exhibit short terminal inverted repeats. The transposase gene encodes a protein that catalyses excision, replication and site specific insertion of the transposable element. Since transposable elements are capable of inserting into chromosomes, plasmids and bacteriophages, they constitute an important role for generating genetic diversity and dispersal of potentially advantageous genes.

1.8.1.1 Insertion Sequences

Insertion sequences (IS) are the simplest variety of transposable elements as they exclusively encode genetic information required for transfer to a new location. IS measure 1-2.5Kb in size and are very common in prokaryotes, where several hundred have been characterised (Madigan *et al.*, 2000a). *S. aureus* typically harbours three types (IS*1181*, IS*431*, and IS*2721*) in addition to others, including remnants (Baba *et*

al., 2008). These IS are often found in multiple copies within the genome, as exemplified by IS*1181* featuring eight and ten times in the N315 and Mu50 genomes, respectively (Kuroda *et al.*, 2001). Although IS do not carry antimicrobial resistance or virulence genes, they play an important role in gene regulation, as demonstrated by involvement of IS*431* in regulation of β-lactam resistance expression (Kobayashi *et al.*, 2001).

1.8.1.2 Transposons

Transposons are larger than IS and, in addition to encoding genes for mobility, they often carry genes for antimicrobial resistance. The most common *S. aureus* transposon is Tn*554* and provides resistance to erythromycin (*ermA)* along with spectinomycin (*spc*) (Phillips and Novick, 1979). A pseudo Tn*554* exists that carries cadmium resistance (*cad*) and interestingly has recombined with Tn*554* to produce a pseudo Tn*554* carrying the *ermA* and *spc* resistance genes (Ito *et al.*, 2001; Ito *et al.*, 2003). After hitchhiking entry into a cell using another MGE, it is common for transposons to duplicate and insert into multiple sites across the new host genome (Baba *et al.*, 2008). Other *S. aureus* transposons include Tn*5801* (tetracycline resistance), Tn*552* (β-lactamase resistance) and a Tn*916*-like transposon (Kuroda *et al.*, 2001; Holden *et al.*, 2004). The most clinically significant transposon found in *S. aureus* is the vancomycin resistance transposon, Tn*1546*, which originated from *Enterococci* (Weigel *et al.*, 2003). It has been suggested that after entrance into a *S. aureus* cell via a conjugative plasmid, Tn*1546* transferred spontaneously into the host chromosome, while the plasmid was not retained (Clewell *et al.*, 1985; Weigel *et al.*, 2003). In addition to directly affecting antibiotic treatment, transposons play an important role in the diversification of the *S. aureus* accessory genome as they facilitate cross species genetic transfer.

1.8.2   Plasmids

Plasmids carried by *S. aureus* are classified into three classes based on their size and ability to conjugate (Paulsen, 1997). Class I plasmids are the smallest in size (typically <5Kb) and include pT181 and pUB110. pT181 encodes resistance to tetracycline and is found integrated within particular SCC*mec* variants and also in free form, where up to 20 copies can be isolated from a single isolate (Novick *et al.*, 1982; Ito *et al.*, 2001). Similarly, pUB110 is associated with selected SCC*mec* elements, however this integrated plasmid carries resistance to tobramycin and kanamycin (*aadD*) and for bleomycin (*ble*) (McKenzie *et al.*, 1986). In addition to the above well-characterised plasmids, small plasmids lacking obvious functions, called cryptic plasmids, have also been identified (Diep *et al.*, 2006).

Class II *S. aureus* plasmids are larger (up to 40kb) and are generally characterised by carrying resistance to one or more of the following antimicrobials: β-lactams (*bla*), heavy metals (*ars*, *cad* or *mer*), antiseptics (*qac*) or aminoglycosides (*aacA-aphD*) (Lindsay and Holden, 2006). The N315 class II plasmid (pN315) contains cadmium resistance, arsenate resistance and a *bla*Z harbouring Tn*554*-like transposon (Kuroda *et al.*, 2001). Interestingly, the Tn*554*-like transposon and arsenate resistance genes were found to have similar organisation in the class II plasmid pI258. As the cadmium resistance gene is more similar to other plasmids than pI258, it is suggested that inter-plasmid recombination lead to construction of pN315 (Kuroda *et al.*, 2001).

Class III *S. aureus* plasmids are the largest of the three classes (up to 60Kb) and share many features of the class II plasmids; however the defining difference is the carriage of transfer (*tra*) genes that mediate conjugation. One example from *S. aureus* is the pMu50 that encodes the conjugation–related nicking enzyme t*ra*A, along with resistance to aminoglycosides and quaternary ammonium compounds (*qacA*) (Kuroda *et al.*, 2001). Comparison to larger conjugative plasmids revealed that pMu50 does not feature a complete *tra* set, and therefore is defective in its conjugation abilities (Kuroda *et al.*, 2001).

S. aureus plasmids can also carry genes for virulence factors, with conspicuous examples being enterotoxin D (SED) carried by pIB485 (Bayles and Iandolo, 1989), and exfoliative toxin B (ETB) that is carried by pRW001 (Jackson and Iandolo, 1986). The variety of advantageous genes relating to bacterial fitness is testament to the role plasmids play during pathogenic and resistance evolution in *S. aureus*.

1.8.3   The Staphylococcal Cassette Chromosomes (SCC)

The recently discovered Staphylococcal Cassette Chromosome (SCC) family is a class of MGEs whose transmission between isolates occurs via an unknown mechanism. Members of the SCC family vary considerably in length, structure and content, however, each SCC shares a precise genomic insertion site and associated inverted and direct repeats (Ito *et al.*, 2001). To date, SCC elements have been identified within *S. sciuri* (Juuti *et al.*, 2005), *S. hominis* (Katayama *et al.*, 2003), *S. epidermidis* (Wisplinghoff *et al.*, 2003), *S. haemolyticus* (Hanssen and Sollid, 2007) and *S. aureus* (Ito and Hiramatsu, 1998). Due to the carriage of the methicillin resistance gene, *mec*A, SCC*mec* elements are the most clinically important SCC. Acquisition of SCC*mec* by *S. aureus* results in the formidable human pathogen MRSA. Therefore, characterisation or genotyping of these elements is of great interest to clinicians and also for global population studies.

1.8.4   The Staphylococcal Cassette Chromosome mec (SCCmec)

The location of the *mec*A gene within the *S. aureus* genome was elucidated during the 1970s. Initially, it was recognised that *mec*A was located chromosomally, as chromosomal rather than plasmid DNA was shown to mediate the transformation of methicillin-sensitive to resistance cells (Sjostrom *et al.*, 1975). Subsequent studies provided evidence that *mec*A was part of an exogenous inserted element, as *mec*A proximal sequence was not discovered within MSSA (Stewart and Rosenblum, 1980; Beck *et al.*, 1986). Cloning and sequencing of *mec*A (Inglis *et al.*, 1988; Matsuhashi *et al.*, 1986; Song *et al.*, 1987) led to detection of the upstream regulatory genes, *mec*R1 and *mec*I (Tesch *et al.*, 1990; Hiramatsu *et al.*, 1992). In addition, the notion

of this variable region being clinically important was strengthened as additional antimicrobial resistance traits (Tn*554* and pUB110) were found in close proximity to *mec*A (Dubin *et al.*, 1991). Ultimately, the entire chromosomal structure featuring *mec*A and its associated exogenous DNA was cloned, sequenced and termed the Staphylococcal Cassette Chromosome *mec* (Ito *et al.*, 1999; Katayama *et al.*, 2000).

The origin of SCC*mec* is unknown, however there is evidence suggesting it originated from an unidentified source, then transferred from coagulase-negative staphylococci (CoNS) to *S. aureus* (Archer *et al.*, 1994; Couto *et al.*, 1996; Katayama *et al.*, 2003; Wu *et al.*, 1996). The SCC*mec* mode of transmission is unknown and has been subject to debate. Regardless, insertion of SCC*mec* into the *S. aureus* genome is an infrequent event, and is thought to have occurred on only 20 occasions (Robinson and Enright, 2003). Overall, the SCC*mec* constitutes a novel class of distinct MGEs that range in length from about 20KB to 52Kb (Chongtrakool *et al.*, 2006). Since the discovery of the first SCC*mec* elements, numerous variants have been characterised based upon three general structural features.

1.8.4.1 The Cassette Chromosome Recombinase (ccr) Complex

The first core SCC*mec* domain hosts the cassette chromosome recombinase (*ccr*) gene(s) and is thus termed the *ccr* complex. The *ccr* complex facilitates the site-specific excision and integration of SCC*mec* into the conserved *orfX* located near the origin of replication (Katayama *et al.*, 2000; Ito *et al.*, 1999). The first three SCC*mec* elements were found to carry different *ccr* complex types that featured allelic variants of two adjacent *ccr* genes (*ccr*A1-3 and *ccr*B1-3) (Ito *et al.*, 2001). The two *ccr* genes were considered essential for mobility until a CA-MRSA SCC*mec* element was recovered with only a single, yet functional, *ccr* gene (*ccr*C or *ccr*5) (Ito *et al.*, 2004). In addition to the primary four *ccr* types, a rare SCC*mec* element was found to carry *ccr*4 (Oliveira *et al.*, 2001). The *ccr* encoded proteins share N-terminal domain homology with site-specific recombinases from the invertase-recombinase family, which are thought to facilitate strand-exchange during recombination (Ito *et al.*,

1999). Although the exact mechanism for SCC*mec* transmission has not been elucidated, the *ccr* genes play a crucial role in converting MSSA to MRSA.

1.8.4.2 The mec Complex

The second standard SCC*mec* component is termed the *mec* gene complex and encompasses the ORFs adjoining *mec*A. The ORFs include *mec*R1 and *mec*I, which are involved in regulating the inducible expression of methicillin resistance. These are often inactivated in clinical MRSA strains, as an intact *mec*I represses *mec*A expression (Suzuki *et al.*, 1993; Kuwahara-Arai *et al.*, 1996). The remainder of the *mec* complex resides downstream of *mec*A and consists of a 2Kb hyper-variable region (HVR) followed by an immobilised insertion sequence termed IS*431mec* (Ito *et al.*, 1999) (Fig 1.2).

The *mec* complex occurs in various forms due to structural variation of the regulatory genes. Currently, three main *mec* complexes are recovered in MRSA: the complete *mec* class A (*mecI-mecR1-mec*A-IS*431mec*), and the truncated *mec* class B (IS*1272*-ΔmecR1-mecA-IS*431mec*) and *mec* class C2 (IS*431*-ΔmecR1-mecA-IS*431mec*) (Katayama *et al.*, 2001; Lim *et al.*, 2003; Ito *et al.*, 2004). Variations of these primary *mec* classes are found within MRSA and also CoNS (Katayama *et al.*, 2001).
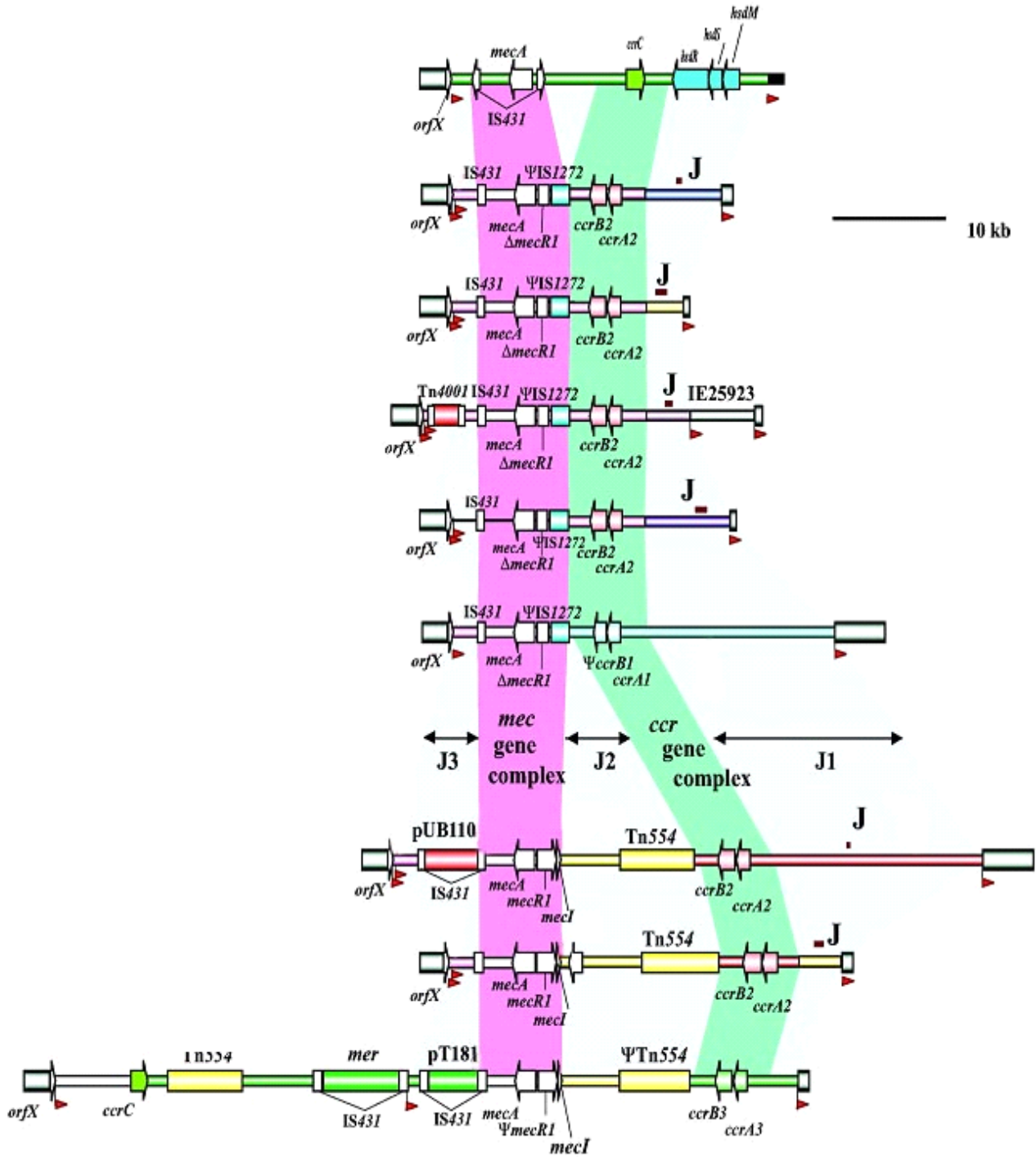
Fig 1.2. Schematic representation of the major SCC*mec* structural types. Each SCC*mec* element integrates into *orfX* and consists of a *mec* class complex, a *ccr* complex allotype and regions of non-essential function termed Junkyards. The SCC*mec* elements vary in length due to integration of MGEs and via carriage of type specific sequence. Numerous SCC*mec* variants exist resulting from MGE deletion/insertion and recombination events. Figure modified from Chongtrakool and co-workers (Chongtrakool *et al.*, 2006)

1.8.4.3 The "Junkyard" Regions

The remaining SCC*mec* sequence not associated with mobility or methicillin resistance is considered non-essential and is therefore referred to as regions of "junkyard" DNA. Three junkyard regions have been classified (J1-3), with locations relative to the positions of the *ccr* and *mec* gene complexes (J1-*ccr*-J2-*mec*-J3) (Ito *et al.*, 2003). Although considered to be of limited or no function, the junkyard DNA attracts and hosts auxiliary genes in the form of antimicrobial resistance plasmids, transposons and insertion sequences, in addition to other features such as a restriction-modification system (Ito *et al.*, 2001; Ito *et al.*, 2004). Junkyard regions are primarily responsible for the observed differences in overall SCC*mec* length as the quantity and type of auxiliary features harboured are considerably diverse. Furthermore, by capturing MGEs, the junkyard regions play an important evolutionary role in diversifying SCC*mec* structural types in a manner analogous to diversification of the core genome backbone by the accessory genome component.

1.8.5   SCCmec Nomenclature

Since the first SCC*mec* element was characterised in 1999, discovery of new variants has been common, which in turn has complicated classification of major types and sub-types. Initially, SCC*mec* nomenclature was based on the combination of *ccr* allotype and *mec*-gene complex that each element carried. This classification strategy allowed characterisation of five major structural types, namely: type I - *ccr* type 1 and *mec* class B; type II - *ccr* type 2 and *mec* class A; type III - *ccr* type III and *mec* class A; type IV - *ccr* type 2 and *mec* class B; and SCC*mec* type V, characterised by *ccr* type 5 and *mec* class C2 (Ito *et al.*, 2001; Ito *et al.*, 2004; Ma *et al.*, 2002) (Fig 1.2). Subtype classification of these major types, including SCC*mec* types IVa, b and c, is possible via identification of unique J1 variations corresponding with particular subtypes. Interrogating additional junkyard regions is also utilised for subtyping SCC*mec* types I, II and III.

As more MRSA isolates were tested, new SCC*mec* variants were discovered, which forced the development of a comprehensive (and more complex) nomenclature system (Chongtrakool *et al.*, 2006). This system is also based on the core *mec* class and *ccr* complexes, however junkyard region variation has also been officially incorporated. For example, SCC*mec* element type IV is proposed to be named 2B.1.1 or IV.1.1, where the 2 relates to the *ccr* allotype, the B relates to the *mec* class, the first digit relates to J1 variation and the second digit corresponds with J2 and J3 variation. This proposed nomenclature system appears appropriate for the SCC*mec* variation noted, however, it is important to note that classification of SCC*mec* elements largely dependents on the resolution and methodology used for characterisation.

1.8.6   *SCCmec Family: Structural Variation*

1.8.6.1 *SCCmec Type I (1B) Variants*

The SCC*mec* 1B.1.1 element spans 34.3Kb, characteristically carries type 1 *ccr* complex with *mec* class B and was discovered in the oldest known MRSA strain from 1960 (Ito *et al.*, 2001). Since antimicrobial chemotherapy was in its infancy during this time, there was little selective pressure to gain multi-resistance. As such, SCC*mec* 1B.1.1 confers resistance only to β-lactams. This lack of additional multi-resistance genes affects both the length of the element and its ability to compete with other SCC*mec* types within modern nosocomial environments. Although 1B.1.1 has disseminated globally in the last five decades, it is not often recovered in comparison with other major SCC*mec* types (Coombs *et al.*, 2004). Furthermore, SCC*mec* 1B.1.1 lost its ability to mobilise from the chromosome due to a frameshift mutation in *ccr*B1 (Ito *et al.*, 2001). Interestingly, in an initial effort to identify the structural boundaries of 1B.1.1, type 2 *ccr* genes were used successfully to precisely excise the element (Ito *et al.*, 2001).

SCC*mec* 1B.1.1 variants are subdivided by variation within the junkyard regions. Variant 1B.1.2 (IA) is clinically identified due to carriage of an IS*431* flanked copy of pUB110 within the J3 region. In comparison with the 1960 type I element, 1B.1.2 was

not found in isolates before 1989 (Oliveira *et al.*, 2001), reinforcing the notion that selective pressures within hospitals allow MRSA to adapt through MGE acquisition. 1B.1.2 also carries a truncated HVR region within the *mec* complex (Oliveira *et al.*, 2001), and although this technically should generate a new *mec* complex allotype, it has been neglected.

The commonly used method for detecting SCC*mec* elements based on *mec* class and *ccr* allotype identifies 1B.1.1, and potentially subtype 1B.1.2, if the user includes detection of pUB110. In addition to 1B.1.2, four SCC*mec* 1B subtypes have been reported that are not differentiated by standard characterisation methods. Of particular interest, the J1 region located upstream of the 1B.1.1 *ccr* complex encodes a gene essentially identical to the plasmin-sensitive surface protein (*pls*) previously identified within MRSA (Oliveira *et al.*, 2001; Ito *et al.*, 2001). *Pls* is a large protein that shares a serine and aspartate (SD) repeat region analogous to the *Clf-Sdr* family, and is thought to play a role in preventing host adhesion during infection (Hilden *et al.*, 1996). The *pls* region was considered characteristic for 1B variants and was therefore used as a genotyping target, however a *pls* negative SCC*mec* 1B element has since been reported (Shore *et al.*, 2005) and has not been named according to the proposed nomenclature (Chongtrakool *et al.*, 2006). In addition, the insertion sequence IS*256* has been identified in the J3 region of 1B.1.2 and also within the J3 region of a "1B.1.2", which although carries the characteristic ΔHVR, lacks pUB110. Furthermore, the Egyptian isolate BK793 carries a 1B variant with an insertion of at least 8Kb of unknown sequence within J3 (Oliveira *et al.*, 2001). These four 1B variants are not subdivided from 1B.1.1 using the most commonly used and accepted SCC*mec* characterisation method, and have not been assigned contemporary names. From this point forward, each SCC*mec* variant not assigned a contemporary name will be named according to its original terminology.

1.8.6.2 SCCmec Type II (2A) Variants

SCC*mec* elements characterised via carriage of *mec* class A and the *ccr* 2 complex are termed SCC*mec* 2A. The first SCC*mec* 2A element discovered (2A.1.1) measured

52Kb in length and was identified within the hospital-acquired pre-MRSA strain, N315 (Ito *et al.*, 1999). Pre-MRSA variants are rarely found within clinical isolates, as most have acquired a *mec*I mutation that permits *mec*A expression. In addition to β-lactam resistance, 2A variants have been found to carry pUB110 and Tn*554*, which facilitate a multiresistant phenotype. Upstream of *mec*A and the *ccr* complex in the J1 region, particular SCC*mec* 2A elements carry a *kdp* operon that encodes a potassium-uptake system, which responds to extracellular osmolarity changes (Sugiura *et al.*, 1992). The carriage of *kdp* in SCC*mec* 2A.1.1 showed the potential for SCC*mec* to disseminate not only antimicrobial multi-resistance, but additional genes that may allow *S. aureus* to survive better in a variety of environmental conditions.

The archetype SCC*mec* 2A.1.1 characteristically encodes an upstream *mec*A region consisting of *kdp*-*ccr*AB2-Tn*554* and a downstream *mec*A region containing HVR-IS*431mec*-pUB110-IS*431*-*dcs* (Oliveira *et al.*, 2001). Numerous and varied SCC*mec* 2A elements, however, have since been characterised. The subtypes classified display variation within the junkyard regions and the *mec* complex locus. SCC*mec* 2A.1.2 is essentially identical to 2A.1.1, the difference resulting from loss of pUB110 from the J3 region (more likely the gain by 2A.1.1) (Cha *et al.*, 2005). SCC*mec* 2A.2 displays variation within the *mec*A upstream region with a unique block of sequence in place of the *kdp* operon (Hisata *et al.*, 2005). An additional 2A variant is characterised by a lack of J3 downstream common sequence (*dcs*) (Cha *et al.*, 2005) and a further variant carries the rare *mec* class F (Shukla *et al.*, 2004b). The most significant discovery of SCC*mec* 2A variation was the identification of five new variants (2A.3.1-5) from Irish hospitals (Shore *et al.*, 2005).

Compared with 2A.1.1, the five novel Irish variants displayed significant structural differences. Foremost was the recombinational replacement of the J1 region (*kdp*) with the much shorter J1 region from SCC*mec* 2B, which effectively reduced the overall length of the variants by approximately 12Kb. Furthermore, the variants 2A.3.1 and 2A.3.4 were found to harbour a novel *mec* complex arrangement (*mec* class A.4), which is characterised by an unusual *mec*I insertion of IS*1182* that results in two separated *mec*I fragments. The sole difference between these two variants is

the loss of pUB110 from 2A.3.4. SCC*mec* 2A.3.3 and 2A.3.5 were also found to carry a novel arrangement of *mec*A associated genes, in this case termed *mec* class A.3. This complex is similar to *mec* class A.4, however, the inserted IS*1182* does not produce two *mec*I fragments, as the upstream IS*1182* junction abuts a copy of Tn*554*. These variants also lack 4Kb of J2 sequence that usually lies between the *mec* complex and Tn*554*, which serves to further decrease the size of these variants. The differentiating feature of the two *mec* complex A.3 variants is carriage of pUB110. The final 2A variant characterised by Shore *et al* is 2A.3.2, which in addition to the 2B J1 region, differs from 2A.1.1 through a lack of Tn*554* within the J2 region. The extensive variation in each of the three junkyard regions and the *mec* complex demonstrates that recombination and insertion events play a significant role in diversification of SCC*mec* 2A. It seems highly likely that future investigation of other diverse MRSA collections will lead to additional 2A variants being identified.

1.8.6.3 SCCmec Type III (3A) Variants

The third SCC*mec* structural type is characterised by carriage of *mec* class A and the type 3 *ccr* complex. The first MRSA carrying a 3A element was reported in 1985 from New Zealand nosocomial-acquired isolate 85⁄2082 (Ito *et al.*, 2001). During that decade, this 3A element established itself as the dominant nosocomial SCC*mec* element. An enhanced ability to thrive within selective nosocomial environments was most likely due to carriage of multiple antimicrobial resistance traits. These traits include; two IS*431* flanked plasmids, pT181 providing tetracycline resistance and pI258 providing mercury resistance; a *mec*A downstream Tn*554* with resistance to erythromycin and spectinomycin; and a ψTn*554* upstream of *mec*A with resistance to cadmium (Ito *et al.*, 2001). However, it was the discovery of an additional pseudo *ccr* complex (termed ψ*ccr*) near Tn*554* that further complicated the already complex 3A element (Ito *et al.*, 2001).

Until recently, the archetype 67Kb 3A element was considered the largest SCC*mec* variant; however, discovery of the pseudo *ccr*-complex ~32Kb upstream of *mec*A sparked investigation into the possibility of conjoined SCC elements. The additional

ψ*ccr* was found to share close homology with the *ccr* complex from the novel SCC*mec* type 5C and was therefore termed *ccr*C (Chongtrakool *et al.*, 2006). Location of the *ccr*C complex is positioned ~5Kb from *orfX*, and, notably, is linked to Tn*554* in an analogous manner as the *ccr*3 complex is to ψTn*554* (Ito *et al.*, 2001). Adding further to the notion of separate SCC elements, the 15bp direct repeat sequences usually associated with the right junction of SCC*mec* were discovered in the *ips* region located between the IS*431* flanked pT181 and pI258 (Ito *et al.*, 2001). The defining experiment that proved the existence of sequentially integrated elements separately excised both smaller elements from the genome (Chongtrakool *et al.*, 2006). The term given to the SCC*mec* element under the accepted nomenclature was 3A.1.1. Likewise, the non-*mec*A element carrying *ccr*C, Tn*554* and pI258, was termed the SCCmercury (SCC*mer*) due to carriage of mercuric chloride resistance (Chongtrakool *et al.*, 2006). SCC*mec* 3A diversity was largely detected before the recent taxonomic separation of 3A.1.1 from SCC*mer*. As a result, a selection of 3A variants are characterised through diversity within the flanking SCC*mer*. As such, it is noted below whenever a SCC*mec* 3A variant is defined by SCC*mer* diversity.

The two most common SCC*mec* 3A variants were previously termed IIIA and IIIB, and are now known as 3A.1.2 and 3A.1.3, respectively. 3A.1.2 characteristically lacks pT181, the flanking pair of IS*431*s and *ips*, all of which are typically located within the J3 region abutting SCC*mer*. Determining the SCC*mec* and SCC*mer* junction is not as straightforward as with the original 3A, as the aforementioned *ips* region that contains the right junction direct repeats is absent. From the available evidence, it appears that SCC*mer* joins SCC*mec* 3A.1.2 at IS*431mec* (Oliveira *et al.*, 2001). SCC*mec* 3A.1.3 also does not carry pT181, however, it does carry the *pls/orfX* junction that makes it very similar to the SCC*mer* negative 3A.1.1 carried by JCSC290 (Chongtrakool *et al.*, 2006). A 32Kb 3A variant not conjoined to SCC*mer* has also been discovered (3A.1.4), and differs from 3A.1.1 by harbouring unique *mec* left extremity polymorphisms (Chongtrakool *et al.*, 2006).

The remaining reported SCC*mec* 3A variants primarily display minor variation, with the exception of SCC*mec* III-inv. This remarkable SCC*mec* 3A.1.1/SCC*mer*–like

element displays a reconfigured structure where an internal ~54Kb fragment located between the *tnp*B genes of each Tn*554* (including the *mec* complex) has inverted (Ito *et al.*, 2003). In addition to this unusual rearrangement, III-inv appears to carry ~13kB of unknown sequence, including an IS*431* copy. Interestingly, the inversion has potentially created two novel SCC elements as *mec* complex type A is now linked with *ccr*C (SCC*mec* 2C); conversely, the *ccr*3 complex is now without a *mec* complex and is associated with pI258 (SCC*mer*) (Fig 1.2). Further 3A variants include; a type III without a SCC*mer* Tn*554* MLS (Oliveira *et al.*, 2001), a type III with Tn*554* MLS upstream and downstream of *mec*A (Ito *et al.*, 2003), and two variants, one encoding *pls* and *dcs* (Deurenberg *et al.*, 2005), and the other characterised only with *dcs* (Aires de Sousa and de Lencastre, 2003). The final SCC*mec* 3A element characterised is essentially identical to type III, however *mec* complex A1 is carried, which is characterised by a 166bp deletion within *mec*R1 (Lim *et al.*, 2002). Even though *mec* complexes are determined by *mec*R1 and *mec*I polymorphisms, *mec* class A1 is not widely recognised.

1.8.6.4 SCCmec Type IV (2B) Variants

Due to a lack of superfluous genes within the Junkyard regions, the commonly identified type 2B (IV) elements are amongst the smallest SCC*mec* discovered to date (Ma *et al.*, 2002). The lack of additional antimicrobial resistance is thought to have resulted from SCC*mec* 2B elements being associated with CA-MRSA, as opposed to HA-MRSA that inhabit selective environments (Ma *et al.*, 2002). Importantly, it is suggested the reduced length allows transmission between cells by inadvertent packaging within bacteriophages (Ma *et al.*, 2002).

The type 2B elements are characterised by carriage of *mec* class B and *ccr* complex 2. The regions exterior to the *mec* and *ccr* complexes exhibit characteristic variation for the 2B subtypes. The variants 2B.1, 2B.2.1, 2B.3.1, 2B.4 and 2B.5 are essentially identical; however they are each defined by unique untranslated sequence blocks of varying length within J1. In contrast to 2B3.1, 2B.3.2 does not carry Tn*4001* within J3, and 2B.N.2 unusually carries pUB110 (Cha *et al.*, 2005; Oliveira and de

Lencastre, 2002). Two additional 2B variants (2B.3.3 and 2B.2.2) are identical to 2B.2.1 and 2B.3.1, respectively, but carry novel sequence of unknown function in the lower *mec*A region (J3), which typically harbours the *dcs* (Shore *et al.*, 2005).

1.8.6.5 SCCmec Type V (5C) Variants

Shortly after the discovery of the novel community associated SCC*mec* 2B, additional non multi-resistant SCC*mec* elements from outside the nosocomial environment were identified. SCC*mec* type 5C was first characterised from an Australian CA-MRSA isolate after it failed to type as a known genotype. Full sequence analysis revealed an unusual 27.6Kb structure consisting of a novel *mec* complex (C2) and novel *ccr* (5) (Ito *et al.*, 2004). The *mec* class C2 is characterised by complete deletion of *mec*I and a 976bp deletion of *mec*R1 resulting from insertion of IS*431*. IS*431* has previously been associated with *mec* complex reconfigurations in CoNS (Katayama *et al.*, 2001). In combination with the novel *mec* class C2, SCC*mec* 5C carries *ccr* type 5. The *ccr* types 1-4 are each composed of two conjoined units (A and B). In contrast, type 5 is composed of a single unit and is orientated in the opposite direction. SCC*mec* mobility is preserved as *ccr*5 facilitates precise excision and integration within the *S. aureus* genome (Ito *et al.*, 2004).

In addition to these novel features, SCC*mec* 5C is distinguished further from the four major structural types by carriage of a restriction-modification system within J1 (Ito *et al.*, 2004). Each *S. aureus* genome sequenced to date carries two incomplete restriction-modification systems, in comparison, the 5C restriction-modification system contains each component (*hsdM*, *S* and *R*). Restriction-modification systems protect the host cell from invasion of foreign DNA, such as from bacteriophage infection. Interestingly, sequence homology of *hsdM* is highest with *Lactococcus sakei*, in contrast to the local homologues (Ito *et al.*, 2004; Kuroda *et al.*, 2001). Of particular note, SCC*mer* carries both *ccr*5 and also a *hsd* remnant. It appears plausible that 5C and SCC*mer* share a common precursor SCC. Overall, the complete 5C restriction-modification system is thought to stabilise the *S. aureus* genome, as

attempts to extract the excised element resulted in destruction of the host genome (Ito *et al.*, 2004).

The characterisation of 5C prompted updating of SCC*mec* genotyping strategies that in turn lead to the discovery of additional 5C variants. The majority of new variants were discovered within Australian CA-MRSA, and displayed novel *mec* and *ccr* complex combinations. For example, SCC*mec* elements carrying *ccr* type 5 were discovered with *mec* class B, B1 and E (O'Brien *et al.*, 2005). The *mec* class B1 does not appear to have a *mec*RI or *mec*I, instead a copy of IS*1272* is located directly upstream of *mec*A. The *mec* class E is similar to A1, but does not encode *mec*I (Lim *et al.*, 2003). In addition to these novel combinations, a SCC*mec* element with *mec* class C2 was found with *ccr* type 2 (O'Brien *et al.*, 2005). This SCC*mec* most likely evolved via conversion of *mec* class A to C2 by insertion of IS*431* into the *mec*R1 region. Although full sequence characterisation of these SCC*mec* types was not undertaken, the results highlight that variation can occur within small and remote *S. aureus* populations.

An investigation of non typable SCC*mec* elements in Taiwan uncovered a novel variant (type VT) that encodes both *ccr* type 2 and a variant of type 5 (*ccr*C2) (Boyle-Vavra *et al.*, 2005). The single *ccr* gene was considered a type 5 variant due to sequence homology of 90.3% with *ccr*5 from SCC*mec* 5C. The identified *mec* complex has been named C2, however it differs from that of 5C due to additional direct repeating units within the HVR and a nonsense mutation within the IS*431* upstream of *mec*A. Even though these are minor differences, each provides useful information about SCC*mec* VT's evolutionary history.

An unusual SCC*mec* variant harbouring *ccr*5 and *ccr*2 was isolated in Switzerland and was characterised by complete sequencing (Heusser *et al.*, 2007). The sequencing revealed that the *ccr*5 complex was located in the J3 region downstream of *mec*A, and the *mec*A region contained a previously unreported configuration. The transposon Tn*4001* that encodes aminoglycoside resistance and is flanked by copies of IS*256* had inserted within *mec*R1, thus creating the novel *mec* class B2. Interestingly, flanking

the upstream fragment of ΔmecR1 was a *hsdR* remnant that shared homology with the corresponding sequence from the MW2 genome. Mosaic SCC*mec* structures that form after multiple SCC insertion events provide evolutionary insights and also provide potential genotyping targets.

1.8.6.6 SCCmec Type VI (4B) Variants

SCC*mec* 4B is a small, rare SCC*mec* element that was isolated primarily in Portugal during the 1990s (Oliveira *et al.*, 2006a). It does not carry antimicrobial resistant genes apart from *mec*A and displays a similar overall structure to SCC*mec* 2B. Coincidentally, the structural characteristics of both 4B and 2B were first described simultaneously as the fourth SCC*mec* type, however, this nomenclature issue has since been rectified (Oliveira *et al.*, 2006a; Chongtrakool *et al.*, 2006). The main difference compared with 2B is the novel *ccr* type 4, which shares moderate homology with *ccr* type 3. SCC*mec* type 4B also harbours a unique J1 region of unknown function. Although rare, the *ccr* type 4 provides an insight into the SCC*mec* transmission to *S. aureus*, as *ccr* type 4 was recently identified in a *S. epidermidis* SCC (Mongkolrattanothai *et al.*, 2004).

1.8.7   Bacteriophages

In *S. aureus*, each sequenced genome carries between one and four prophages (Iandolo *et al.*, 2002; Baba *et al.*, 2008). The prophages are classified into five family types (ΦSa1-5), which are defined by integrase gene homology. As the integrase gene sequence dictates the insertion site, it is postulated that multiple phages from the same family cannot coexist within a single isolate (Lindsay and Holden, 2004).

Genetic differences between closely related phages contribute to bacterial strain-to-strain differences, even between closely related isolates. Evidence for this was found in the Mu50 and N315 genome sequences where two and one phage(s) were identified, respectively (Kuroda *et al.*, 2001). Mu50 uniquely carries a prophage from the ΦSa1 family designated ΦMu50B. Interestingly, this prophage is located at the

*att*B site for ΦETA, however in comparison with ΦETA, ΦMu50B does not carry exfoliative toxin A and only exhibits an ORFs homology of 29% (Yamaguchi *et al.*, 2000). In addition to sequence blocks from ΦETA, ΦMu50B contains ORFs with similarity to Φ11 and 80α, which supports the evolutionary hypothesis of module exchange (Kuroda *et al.*, 2001).

Prophages from the ΦSa2 family have been discovered in numerous *S. aureus* genomes. Of clinical importance, the ΦSa2 phages in MW2 and USA300 carry the Panton-Valentine leukocidin (*pvl*) components, *lukF-PV* and *lukS-PV* (Baba *et al.*, 2002; Diep *et al.*, 2006). The *pvl* toxin has an important clinical association, as correlations with the potentially fatal necrotising pneumonia have been identified (Labandeira-Rey *et al.*, 2007). Both phages share a very high sequence and module similarity and reside in community-acquired hosts. The ΦSa2 prophage identified in MRSA252 shares structural and sequence similarity with ΦSa2 prophages from MW2 and USA300, however, ΦSa2 MRSA252 does not carry *pvl* or any other virulence genes (Holden *et al.*, 2004).

Prophages from the ΦSa3 family are found in each *S. aureus* genome sequence, with the exception of COL (Lindsay and Holden, 2006). The ΦSa3 *attB* lies within the β-haemolysin virulence gene that is deactivated after prophage insertion (Smeltzer *et al.*, 1994). The ΦSa3 phages from the closely related Mu50 and N315 share 77% amino acid homology and encode different enterotoxin alleles (Kuroda *et al.*, 2001). Additionally, the corresponding phage in the unrelated MW2 carries two additional enterotoxin genes (Baba *et al.*, 2002). Each ΦSa3 was also found to carry staphylokinase A thrombolytic enzyme (*sak*), which is postulated to assist bacterial spread within a host due to its fibrin-specific, blood clot dissolving properties (Collen, 1998). A subset of the ΦSa3 phages carry the chemotaxis inhibiting protein, which is an anti-inflammatory agent that inhibits C5a-dependent recruitment of neutrophils (de Haas *et al.*, 2004), and most ΦSa3 prophages encode enterotoxin A (*sea*) (Kuroda *et al.*, 2001). Alignment of ΦSa3 prophage sequences from different genomes show instances of multiple cross-over events leading to mosaic structures (Baba *et al.*, 2002).

The ΦSa4 and ΦSa5 prophage families are not found extensively in *S. aureus*. MSSA476 is the only sequenced *S. aureus* genome that hosts a prophage from the ΦSa4 family (Holden *et al.*, 2004). The MSSA476 ΦSa4 integrase gene resembles that of phage L54a from the COL genome and inserts within the promoter of the putative serine protease termed *htrA* (Holden *et al.*, 2004). Unlike many of the other prophages that infect *S. aureus*, no virulence genes have been identified. However, MSSA476 ΦSa4 presents a mosaic structure with module similarities to the ΦSa2 prophages from MRSA252 and MW2, in addition to Φ12 (Iandolo *et al.*, 2002; Holden *et al.*, 2004). Lastly, a ΦSa5 prophage has only been identified in the genome of NCTC8325, where it is termed Φ11. This prophage does not carry any virulence or drug resistance genes (Lindsay and Holden, 2004).

Bacteriophages play an important role in augmenting the *S. aureus* accessory genome as they transmit virulence factors and can inadvertently facilitate transfer of random DNA between strains and species via generalised transduction. Although the newly acquired factors are superfluous for basic survival, it has been postulated that imported prophage traits can lead to phenotypic change and adaptation of pathogens to new hosts, in addition to evolution of new pathogens and epidemic clones (Brussow *et al.*, 2004). Overall, the impact of phages on bacterial genome evolution is significant, and may benefit genotyping strategies as the associated variation is often unlinked to the core type.

1.8.8   Genomic Islands

S. aureus genomic islands are large (~40Kb), non-mobile elements found in identical locations across all genomes sequenced to date (Baba *et al.*, 2008). Two types of genomic islands termed υSaα and υSaβ exist and carry numerous pathogenicity genes (Kuroda *et al.*, 2001; Holden *et al.*, 2004). Genomic island origins and mechanism of mobility are unclear, however horizontal transfer appears to have occurred.

The υSaα genomic islands are characterised primarily by carriage of superantigen exotoxins (*set*) and lipoprotein (*lpl*) clusters (Williams *et al.*, 2000; Baba *et al.*, 2002). The general structure of υSaα is conserved; however allelic variation is evident. An examination of the exotoxin cluster from phylogenetically distinct strains showed that gene loss and recombination significantly contributed to observed evolutionary divergence (Fitzgerald *et al.*, 2003). This is also exemplified by the υSaα from N315 carrying one more *set* gene than the closely related Mu50 (Kuroda *et al.*, 2001). Interestingly, comparison of the orthologous *set* genes revealed that the majority of differences were located within the T-cell receptor binding domain (Papageorgiou and Acharya, 2000; Baba *et al.*, 2002). This accumulation of subtly different exotoxins has lead researchers to label such clusters as a "nursery", which has the capacity to rapidly produce a battery of toxins when required (Jarraud *et al.*, 2001). In addition to the exotoxin set, each υSaα element harbours 3 - 9 lipoprotein genes that are found in allelic forms analogous to the upstream *set* cluster. Other genes typically carried by υSaα variants are a degenerated *tnp* gene and a restriction-modification system composed of *hsd*S and *hsd*M, which appears to provide genome stabilisation for the genomic island (Kuroda *et al.*, 2001; Baba *et al.*, 2002).

The υSaβ genomic islands are similar in length and overall structure compared with υSaα, however distinct differences in the carriage of virulence genes has been observed (Baba *et al.*, 2002). In contrast to υSaα, υSaβ allelic variants carry clusters of superantigen enterotoxin genes and serine proteases, in addition to leukocidin genes (*luk*D and *luk*E). Like υSaα, the υSaβ also sub-types into allelic forms based on gene variation. Instead of an enterotoxin cluster, the υSaβ allelic variants carried by MW2 and NCTC8325 carry a novel gene cluster consisting of bacteriocin genes and their modification factors (Baba *et al.*, 2002). These genes have homology to some reported in *S. epidermidis* (Schnell *et al.*, 1988) and are thought to provide a competitive advantage against the host flora. As with υSaα, υSaβ carry *tnp* genes that encode a frameshift mutation, and a defective restriction-modification system (Baba *et al.*, 2002).

The *hsd*S genes carried by υSaα and υSaβ feature a mutual association with the overall structural type and are therefore used to define each allotype accordingly (Baba *et al.*, 2002). N315 and Mu50 carry type I υSaα and type I υSaβ, NCTC8325 and COL carry type I υSaα and type II υSaβ, whereas MW2 carries type II of both islands and MRSA252 carries type III allotypes (Baba *et al.*, 2002; Baba *et al.*, 2008). Even though the *S. aureus* genomic islands are not actively mobile, they provide the host with a battery of toxins and appear capable of evolving to suit the host and their environment.

1.8.9   *Pathogenicity Islands*

S. aureus pathogenicity islands (SaPI) are 14-18kb mobile genetic elements that typically encode superantigens and additional virulence genes (Novick and Subedi, 2007). Superantigens are molecules that activate large populations of T-cells by forming ligands with MHC class-II molecules that ligate with the Vβ portion of the T-cell receptor (Herman *et al.*, 1991). This non-specific reaction activates from 10-50% of all T-cells that subsequently stimulate a cytokine release that generates symptoms associated with superantigen shock (Choi *et al.*, 1990). The most prominent Staphylococcal superantigen is the toxic shock syndrome toxin-1 (TSST-1) encoded by the *tst* gene, which is responsible for severe menstrual and non-menstrual toxic-shock syndrome (Bergdoll *et al.*, 1981). Although approximately 20% of all *S. aureus* have been estimated to carry TSST-1, carriage is restricted to a few select lineages (Moore and Lindsay, 2001; Peacock *et al.*, 2002). In addition to *tst*, SaPI often carry superantigens responsible for food poisoning and include numerous enterotoxins, including B, C, K and L (Kuroda *et al.*, 2001; Baba *et al.*, 2002). High virulence potential of the superantigens makes the SaPI important components of the accessory genome.

SaPI have been identified in nearly all available genome sequences and more than 20 SaPIs have been sequenced (Novick and Subedi, 2007). Common features include an integrase gene, a *Rep* protein, a terminase small subunit homolog and flanking direct repeats. The majority of remaining SaPI ORFs do not share close sequence homology

with any other gene identified to date (Novick and Subedi, 2007). Classification of the known SaPIs into five subgroups is based on the *att* sequences that dictate the genome insertion point (Novick and Subedi, 2007). As has been demonstrated in bacteriophages, there are no instances where a single strain carries more than one copy of each type. As each genome provides five insertion sites, the SaPIs have the potential to influence pathogenicity and to generate genetic diversity within the species.

SaPIs are transferred horizontally between genomes via the actions of a "helper" phage (Lindsay *et al.*, 1998). During manufacture of the helper bacteriophage, the SaPIs are excised and inadvertently packaged into particles instead of the phage DNA. The SaPI "hijacked" phage particle encases a ~15Kb genome and therefore displays a smaller head in comparison with the particles from the parent phage that encase a genome of ~40Kb (Ruzin *et al.*, 2001). After bacteriophage infection of the new host, the SaPI integrates itself into the host chromosome in both a site and orientation specific fashion. This transfer is additionally beneficial to the host cell as the defective SaPI hijacked phage cannot lyse the cell. This unregulated process of genetic transfer is thought to occur frequently in nature and appears to be the only feasible mechanism for non-phage related genetic elements to transfer horizontally within *S. aureus* (Lindsay and Holden, 2006).

1.9   *S. AUREUS* GENOTYPING

Bacterial typing is the act of subdividing bacteria into smaller groups that share observable traits. Typing using the traditional phenotypic methods such as bacteriophage typing (Blair and Carr, 1960), have largely been replaced by powerful molecular genotyping methods. Genotyping characterises bacteria by the production of unique genetic "fingerprints" that are recognisably different among strains of the same species (Maslow *et al.*, 1993). These characteristic fingerprints are generated by using common methods such as PCR amplification, enzyme digestion, electrophoresis, hybridisation and DNA sequencing techniques.

The level of typing discrimination attained among strains is dependent upon the method used, and the diversity within the collection of isolates tested. There are two general types of genotyping methods: comparative, and library. Comparative genotyping methods determine whether isolates are the same or different and are useful for either preventing or managing outbreaks by distinguishing between background and infecting isolates (Tenover *et al.*, 1997). Library methods demonstrate extended capabilities, as they report the genetic changes that define the genotype. This feature is essential when reconstructing population structures, from micro-environments (such as hospitals), to large-scale global surveys (Deshpande *et al.*, 2004; Montesinos *et al.*, 2002). To date, a standard MRSA genotyping protocol has yet to be implemented globally. Within the last decade, numerous genotyping methods have been devised to accommodate this need.

1.9.1 *Restriction Digestion Methods*

1.9.1.1 *Pulsed-Field Gel Electrophoresis (PFGE)*

Pulse-Field gel electrophoresis (PFGE) was a method developed in 1984 to separate and display yeast chromosomes for the purpose of karyotyping (Schwartz and Cantor, 1984). Unlike conventional electrophoresis that is limited to separating fragments of less than 50Kb using a constant and unidirectional electric field, the PFGE apparatus alters the electric field orientation after each pulse (Chu *et al.*, 1986). The programmed reorientation of the field forces the DNA fragments to frequently realign their migratory path. The time taken to realign determines the overall migration distance; the larger fragments reorientate at a slower rate and therefore do not progress along the field as far as smaller fragments during each pulse. PFGE has become a particularly important tool for bacterial genotyping.

Following successful PFGE macro-restriction profiling of a digested *Escherichia coli* chromosome in 1987 (Smith *et al.*, 1987), the potential application of PFGE for bacterial genotyping was realised. Suitability of PFGE is attributed to this technique's ability to generate restriction patterns that differentiate isolates at high-resolution, whilst, importantly, retaining similar patterns between closely related isolates. The high level of differentiation is primarily a result of the number and location of

restriction sites in the genome. This constitutes an advantage for PFGE as it is sensitive to restriction site variation in both the *S. aureus* core and accessory genome components. Furthermore, the MGEs within the accessory genome contribute to resolution as insertion or deletion events alter the restriction fragment lengths (Kuroda *et al.*, 2001; Tenover *et al.*, 1995). As these evolutionary events are somewhat frequent, PFGE is ideally suited for local or short-term investigations where epidemiologically linked isolates are required to be distinguished from sporadic or background isolates. The causative strains from numerous hospital and community *S. aureus* outbreaks have been genotyped successfully in this fashion (Lin *et al.*, 2007; Otter *et al.*, 2007; Gilbert *et al.*, 2006; McDougal *et al.*, 2003). As such, PFGE has been regarded as the "gold standard" method for *S. aureus* genotyping (Bannerman *et al.*, 1995; Chung *et al.*, 2000).

Notwithstanding the regard in which PFGE is held as a method for *S. aureus* genotyping, it does have limitations. In addition to being laborious, time consuming and expensive, the major shortcoming of PFGE is the significant alteration of restriction patterns in response to minor genomic changes (Tenover *et al.*, 1994). This sensitivity creates an overabundance of restriction patterns that are difficult to interpret, serving to complicate the genotyping of long-term and diverse collections since strain relatedness and clonal ancestry are obscured (Blanc *et al.*, 2002). Furthermore, the complex PFGE restriction patterns in combination with the gel based output results in inconsistency among typing laboratories in regard to interpretation of restriction patterns (Cookson *et al.*, 1996; van Belkum *et al.*, 1998b). Although computer software has been developed to assist with the analysis and storage of restriction profiles, interpretation difficulty and a degree of subjectivity often remains (van Belkum *et al.*, 1998b). Due to these inter-laboratory reproducibility issues, Tenover and co-workers devised a standardised protocol for interpreting PFGE restriction patterns (Tenover *et al.*, 1995). The scheme for deducing relatedness among strains is based on the following criteria: if a PFGE restriction pattern differs from the outbreak control by two or three bands then the isolates are *probably* related, differing by four-six bands means they are *possibly* related, and differing by more than seven bands means they are *probably unrelated* (Tenover *et al.*, 1995). Although

effective, it was stressed that this criteria is only applicable in situations where diversity is restricted, such as in short-term outbreaks.

In spite of the reported limitations, PFGE has been used successfully for long-term and large scale population analyses of *S. aureus*. The results of numerous studies in Central Europe and the UK resulted in classification of over 20 PFGE clones, the majority were named after their predominant geographical location (Blanc *et al.*, 1999). A study by Oliveira and co-workers genotyped 3,067 isolates from Southern and Eastern Europe, Latin America and the United States and discovered that 70% could be assigned to just five major lineages (Oliveira *et al.*, 2001). More recently, in an attempt to establish a standardised US national database of PFGE profiles, 957 isolates were genotyped, of which 68% were associated into eight major lineages (USA100-800) (McDougal *et al.*, 2003). Overall, PFGE is a powerful genotyping method, but due to significant drawbacks including; excessive resolving power and complexity of results, PFGE faces competition from a range of novel methods. These methods aim to be rapid, cost effective and function at an appropriate level of resolution so as to determine the relationships among strains from both short and long time periods.

1.9.1.2 Amplified Fragment Length Polymorphism (AFLP)

Amplified fragment polymorphism (AFLP) is another broadly applicable genotyping method that creates highly discriminatory DNA fingerprints (Vos *et al.*, 1995; van den Braak *et al.*, 2004). AFLP generates DNA profiles by selective PCR amplification of restriction fragments followed by electrophoresis. The technique involves three steps. Firstly, total genomic DNA is digested using one frequent and one rare cutting enzyme, followed by ligation of oligonucleotide linkers. This dual enzyme approach allows control over the total number of fragments to be amplified whilst the linker serves as the PCR primer attachment site. The use of multiple enzymes is not crucial as an AFLP technique using only a single enzyme has been reported (Velappan *et al.*, 2001). The second step involves amplification, under stringent conditions, of between 50 and 100 restriction fragments using fluorescent labelled primers. Restricting

amplification to a subset of fragments is achieved by incorporating up to three arbitrary nucleotides to the 3' extremity of each primer. The third step involves denaturing polyacrylamide gel electrophoresis using an automated fluorescent DNA sequencer and analysis of the resulting banding patterns. Sophisticated computer-based techniques for interpretation and analyses of the banding patterns have been developed and are used to define relatedness among isolates (Jarraud *et al.*, 2002).

The first study using AFLP to genotype Staphylococci was conducted on a small collection of isolates and demonstrated limited results; however genotyping to the species and strain level was achieved (Velappan *et al.*, 2001). Subsequent to this, studies that genotyped large collections of *S. aureus* have been conducted. One such study genotyped 198 isolates to determine whether a correlation existed between genetic backgrounds, toxin genes and toxin gene regulation (Jarraud *et al.*, 2002). Using the aforementioned computer analysis of the banding patterns, three discrete groups were resolved: AF1, AF2 and AF3. The classification of these broad groups serves to generalise the actual variation that exists within the collection. If required, higher resolution could have been attained simply by increasing the stringency of criteria for what constituted identical banding patterns (Velappan *et al.*, 2001). A similar yet more recent study utilised high-throughput AFLP to genotype 391 American and 829 Dutch nasal carriage isolates (Melles *et al.*, 2008). This study aimed to determine whether an overlap of genetic backgrounds was evident in non-clinical isolates from different geographical locations. Overall, four comparable genotypes were resolved from each population, with the proportion of isolates per genotype differing among the populations. This AFLP analysis showed that most carriage isolates belonged to a few prominent clones.

In comparison with PGFE, AFLP produces similar resolution and patterns of relationship; however the method offers distinct advantages. A direct comparison between the resolution of AFPL and PFGE revealed comparable discriminatory power and a similar clustering of related strains (Melles *et al.*, 2007). This study also demonstrated better subtype resolution by increasing the stringency for type differentiation arbitrarily. Results for PFGE and AFLP are gel based images, however the digitised AFLP results are considered more amenable for inter-laboratory transfer

and detailed data analysis (Melles *et al.*, 2007). Furthermore, AFLP offers better reproducibility than PFGE, high-throughput automation, faster experimental turnaround time, and the method is available as a standardised commercial kit. AFLP is not without disadvantages, these include reliance upon expensive specialised equipment and complex post run analysis. The main disadvantage of AFLP, however, is that the genetic basis of variation among strains cannot be easily elucidated. This is also a drawback of PFGE and other restriction based methods. Without a comprehensive post reaction sequencing step, the high resolution fingerprints cannot elaborate strain differences beyond that of anonymous banding patterns on a gel. Ideally, a genotyping method would describe the genetic differences between isolates accurately, in the context of a defined population framework.

1.9.2   *Multi-locus Sequence Typing (MLST)*

Multi-locus sequence typing (MLST) was first reported in 1998 as a method for unambiguously characterising bacterial isolates (Maiden *et al.*, 1998). The *S. aureus* MLST scheme is based on PCR amplification and sequencing of seven internal fragments from widely dispersed core housekeeping genes (Table 1.2). As housekeeping genes are essential for cell viability, they are ubiquitous within members of a species and exhibit a slow rate of SNP accumulation. For clonal bacteria such as *S. aureus*, the mutation rate allows MLST to excel in the step-wise reconstruction of evolutionary relationships within the species. The utilisation of an interactive website (www.mlst.net) for data storage and comparative analysis is fundamental to the success of MLST as a modern genotyping method.

Table 1.2. The seven *S. aureus* MLST housekeeping genes.

MLST locus	Gene product	No. of alleles*	Fragment length (bp)
arcC	carbamate kinase	139	456
aroE	shikimate dehydrogenase	178	456
glpF	glycerol kinase	152	465
gmk	guanylate kinase	96	429
pta	phosphate acetyl transferase	134	474

tpi	triosephosphate isomerase	139	402
yqiL	acetylcoenzyme A acetyltransferase	123	516

* Number of alleles as of March, 2008.

The process of generating a MLST Sequence Type (ST) is based around sequencing each housekeeping gene fragment, then using online database search tools to identify a perfect match. Distinct sequences at each locus are known as alleles and are numbered in the database chronologically. The combination of all seven allelic numbers is termed the allelic profile, for example 6, 64, 44, 2, 43, 55, 51, which corresponds with a specific numerical ST, in this case ST93. Novel allelic sequences should be added to the dynamic database, along with associated phenotypic information. Because MLST sequencing results are converted to digital integers and encouraged to be stored online, an indirect global collaboration is formed that aims to assemble comprehensive bacterial population data sets.

Once the tested ST is known, it is possible to determine the relationship of the ST compared with the population framework defined by the database. This is achieved by using the freely available online program, eBURST, which was developed to delineate and visualise evolutionary relationships resulting from core genome SNPs (Feil *et al.*, 2004). eBURST clusters closely related STs into Clonal Complexes (CC) that consist of a progenitor ST and the STs that differ at one and two of the seven alleles, known as single locus variants (SLV) and double locus variants (DLV), respectively. The ability to determine the precise genetic difference between strains is an important attribute of MLST that is ideal for long-term and large scale population studies; as such, MLST schemes for over 40 bacterial species have been established. The *S. aureus* MLST database was established in 1999 and currently features 1057 STs from across the globe. The size of the database reflects the numerous studies that have utilised MLST for both short-term and long-term genotyping applications. The first large scale MLST investigation of *S. aureus* was conducted by Enright and co-workers who genotyped 155 *S. aureus* isolates from community and hospital infections. The resulting 53 STs indicated that MLST for *S. aureus* is highly discriminatory; furthermore, pairs of isolates with identical STs were shown to

display similar PFGE patterns, which served to validate the method. MLST has also been utilised for analysis of the five major pandemic clones that were previously named according to PFGE patterns. MLST results revealed that isolates belonging to the Iberian/archaic clone belong to the CC250, clone V isolates belong to CC8, the Brazilian/Hungarian isolates belong to CC239, the New York/Japan isolates belong to CC5 as was the Paediatric clone isolates (Oliveira *et al.*, 2001). Overall, MLST is a powerful method for genotyping *S. aureus* that adequately discriminates the major hospital acquired clones.

Uptake of MLST as a popular genotyping method coincided with the rise in awareness of MRSA isolated from the general community. To produce a complete epidemiological picture of MRSA globally, studies focussing on CA-MRSA have been conducted and resulted in the identification of common clones. The notorious strain responsible for the deaths of four America children is known as ST1 (Anonymous, 1999), and has also been found to cause community acquired infections in Australia (Coombs *et al.*, 2004). ST30 has been identified within the South Pacific islands, New Zealand, Australia, Singapore and Europe. ST59 has been found in Australia, America and Taiwan. ST80 has been isolated in Germany and ST93 is exclusively found in Australia, and is closely associated with the Aboriginal population (www.mlst.net). The MLST eBURST reconstruction of the *S. aureus* population is shown in Fig 1.3.

Compared with current genotyping methods, MLST has both advantages and disadvantages. The major benefit is production of digital sequence data that allows for unambiguous ST comparison using the internet. The current high standard of database management serves to increases the usefulness of this epidemiological tool. Furthermore, since MLST is DNA sequenced-based it is a simple matter for laboratories to conduct the standardised reactions without the use of reference isolates for each important clone, unlike PFGE (Enright and Spratt, 1999). One disadvantage of MLST is that only polymorphisms within the core genome are surveyed. This is of particular concern for *S. aureus*, as the accessory component of the genome is responsible for the diverse clinical manifestations, including the observed variation in

virulence and antibiotic resistance. This is also of concern for short-term epidemiological investigations where the outbreak strain may not be differentiated from the sporadic isolates. This can be alleviated by using MLST in combination with methods that target the accessory genome (Keim *et al.*, 2004; Enright *et al.*, 2002; Bartels *et al.*, 2007). Even though DNA sequencing is the major strength of MLST, it is also this technique's major drawback. The time and considerable cost involved is not amenable to routine high throughput analysis in clinical laboratories. This shortfall has spawned alternative methods for providing genotype information based on MLST mutations, without reliance on DNA sequencing.
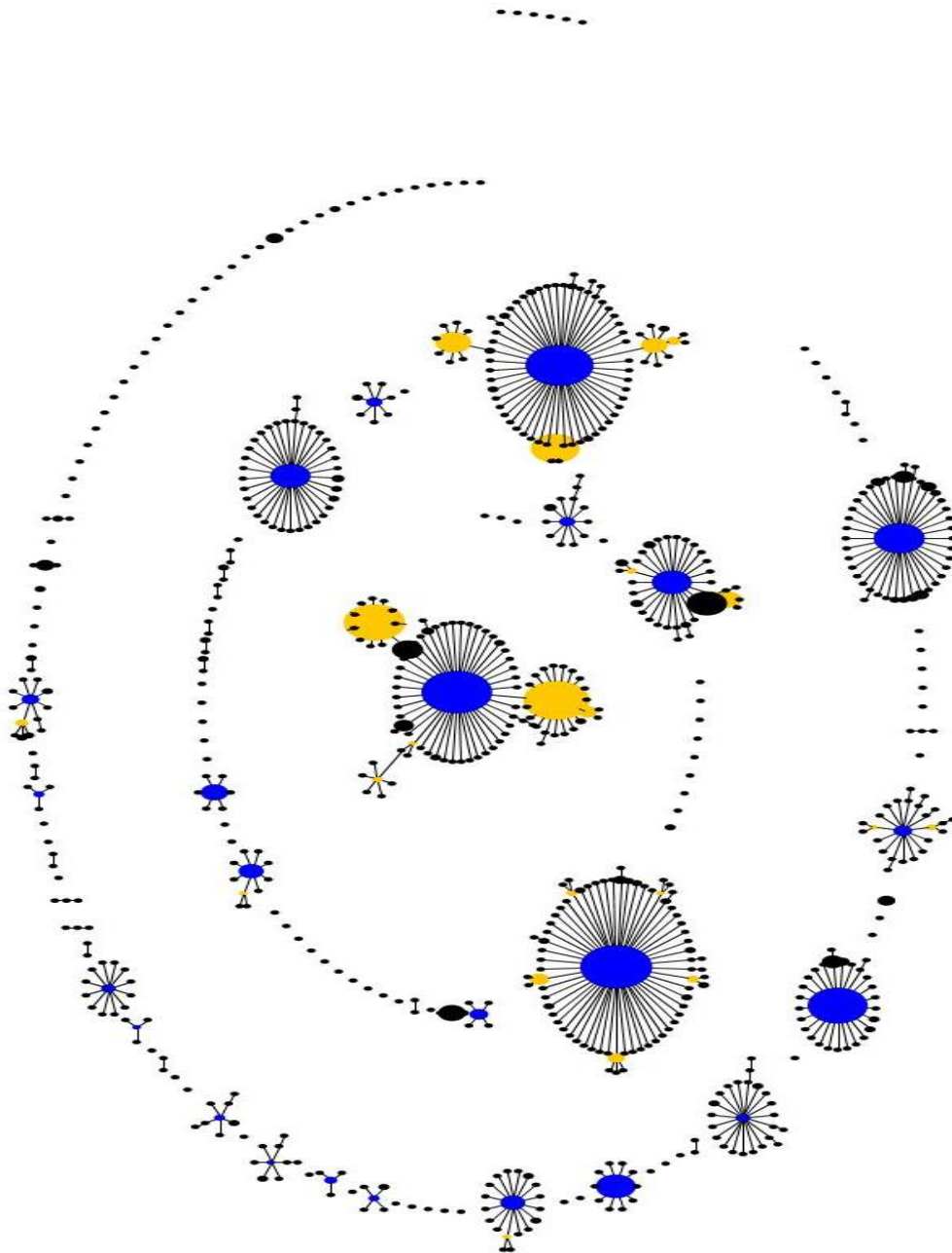
Fig 1.3. The *S. aureus* MLST population structure defined by eBURST. *S. aureus* is primarily a clonal evolving organism, therefore the majority of new strains descend vertically from progenitor strains and produce tight clonal complexes. In this figure, each dot represents a unique ST. The blue dots are the progenitor STs, the yellow dots are the progenitor STs for successful subtypes and the black dots are either closely related STs (linked to progenitor STs) or unlinked singletons. The dot size indicates the relative abundance of that particular ST in comparison to the other STs in the data set. Overall, MLST and eBURST provides an unprecedented visual representation of the *S. aureus* population structure.

As the fragments sequenced for MLST are characteristic of the core genome, the data has been used to formulate inferences regarding genome evolution and structure. Feil and co-workers analysed the MLST data from 334 isolates and reported two main findings (Feil *et al.*, 2003). The first confirmed that the population structure of *S. aureus* is indeed clonal. This confirmation was achieved by observing that the core genome evolves by point mutation 15 times more often than by recombination. The low recombination rate was corroborated in a separate study of large fragment recombination events in *S. aureus* (Robinson and Enright, 2004a). The second main finding was that there is no association between ST and propensity to cause disease. This indicates that virulence potential is not linked to the core genome but is greatly influenced by the accessory genome.

1.9.2.1 MLST Based Genotyping Methods

Considerable effort has been expended on developing methods that genotype bacterial isolates into MLST STs and CCs, without the drawbacks associated with DNA sequencing. One such method for *S. aureus* analyses the SNPs within the PCR products from the seven MLST loci using restriction digests and denaturing gradient gel electrophoresis (DGGE) (Gurtler *et al.*, 2002). The DGGE-MLST approach is reported to be rapid and inexpensive, however due to the requirement for specialised equipment and complicated post-run analysis, the method has not been used widely. A similar approach for *S. aureus* termed Multi-locus restriction fragment typing (MLRFT) is based on amplification of the seven loci, followed by restriction digestion, then conventional gel electrophoresis (Diep *et al.*, 2003). The results are relatively simple and inexpensive to achieve and the genotypes correlate well with PFGE, however, like DGGE-MLST this method is seldom used in place of PFGE or MLST. In addition to these restriction based methods, modern techniques have also been used to genotype *S. aureus* into the MLST framework.

Perhaps the most sophisticated approach for non-sequenced based MLST is use of a specialised DNA microarray (van Leeuwen *et al.*, 2003). The Affymetrix GeneChip microarray features four differential probes for each SNP within the seven MLST loci

available at the time. After the MLST loci are amplified and fluorescently labelled they are hybridised to probes where the signal intensity is recorded. The signal intensities are then cross referenced to the mutation database to determine the ST. The MLST microarray performed adequately between two centres and was successful at identifying 28 and 29 of 30 isolates correctly. Drawbacks of this method are its relative expense, requirement for specialised equipment, complicated data analysis and a mutation set that requires updating as the MLST database expands.

Genotyping methods based on the empirical selection of highly informative SNPs within MLST genes have also been developed. Robertson and co-workers describe a bioinformatics driven approach for selecting combinations of SNPs from MLST databases that together differentiate isolates within species based on Simpson's index of diversity (Robertson *et al.*, 2004). This was achieved using specialised software known as Minimum SNPs. The feasibility of real-time PCR reactions for interrogating the allelic state of the chosen SNPs was also reported. A seven SNP set chosen for *S. aureus* was able to discriminate the STs in the MLST database to a *D* of 0.95, meaning a 95% probability that two randomly selected STs would be called different. The "high *D* SNP" set showed considerable utility, however no detail was reported in relation to concordance with the *S. aureus* population structure. The report also demonstrated the use of the Minimum SNPs for identifying SNPs from other bacterial populations including; *Helicobacter pylori*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Campylobacter jejuni*, *Streptococcus pyogenes*, *Haemophilus influenzae*, *Burkholderia pseudomallei/Burkholderia mallei* and *Enterococcus faecium*. The Minimum SNPs method proved useful at identifying highly efficient combinations of MLST genotyping SNPs that are applicable to the entire population, in contrast to SNPs that are simply genotype specific.

A major conclusion from the study was that SNP based genotyping can never equal the resolution attained using MLST. This is due to the fact that SNP typing methods only survey a small proportion of SNPs that define a complete MLST population. The approach of augmenting SNP typing with higher resolution genotyping methods that survey different regions would significantly increase the overall resolution. This strategy was formalised by Keim and co-workers who describe 'progressive

hierarchical resolving assays using nucleic acids' (PHRANA), which generates high resolution genotypes via analysis of multiple loci that exhibit mutation rates of varying speed (Keim *et al.*, 2004). This notion was tested on *Bacillus anthracis* and produced results that demonstrated that slowly evolving clonal SNPs in combination with hypervariable VNTR regions do indeed produce deeply rooted, yet highly informative genotypes.

1.9.3 Genotyping by Repetitive DNA Analysis

Both eukaryotic and prokaryotic genomes encode numerous and diverse regions of DNA hyper-variability. The repeat regions, particularly of bacteria, are often associated with MSCRAMMs and other cell surface proteins that project binding domains into the extracellular environment. DNA sequence hyper-variability occurs due to slip strand mispairing during cell division; however these regions do not require explicit conservation, as their role accommodates variation in amino acid sequence and length. This characteristic creates disparity between the evolutionary mutation rates of the repeat regions and the core genome, so much so that isolates from the identical genetic lineage can be differentiated (Keim *et al.*, 2004). For this reason, hyper-variable regions are used commonly for bacterial genotyping, and those pertinent for *S. aureus* are discussed below.

1.9.3.1 Staphylococcal Protein A (spa) Sequencing

The most successful *S. aureus* VNTR genotyping method is the sequence analysis of the hyper-variable region X located within the Staphylococcal protein A (*spa*) gene. The *spa* protein is one of the largest *S. aureus* proteins and is responsible for binding immunoglobulin G to allow evasion of the host immune system (Uhlen *et al.*, 1984). The *spa* repeat region was first sequenced for genotyping purposes in 1996 by Frenay and co-workers, who identified between 3 and 15 contiguous repeats and 25 different repeating units (Frenay *et al.*, 1996). The *spa* genotyping results and phage typing generally were concordant, however, *spa* typing demonstrated resolution sufficient to enable subtyping of the phage types.

Sequence typing of the *spa* VNTR did not become widespread until the early 2000s, and is now conducted globally. This success is in part due to construction of an online database (http://spaserver.ridom.de/) that stores *spa* sequence type and individual repeat unit information. From this database, it can be determined that the *spa* repeat region features between 1 and 23 repeat units of 24bp, which themselves display high diversity. In total, 155 different repeat units have been identified, forming a total of 2331 *spa* types. Also contributing to *spa* typing success is the existence of dedicated sequence analysis software that translates sequence chromatograms into *spa* types (Harmsen *et al.*, 2003).

Spa sequence typing used on its own has been shown to accurately predict the clonal lineage of *S. aureus* (Koreen *et al.*, 2004). However, analysis of the *spa* VNTR region is optimally conducted in combination with a clonal genotyping method such as MLST (Keim *et al.*, 2004). As MLST and *spa* typing are both sequenced based, it is increasingly common for genotyping studies to use the methods together to generate comprehensive genotypes (Wagenlehner *et al.*, 2007). Like MLST, *spa* typing produces reproducible and unambiguous sequence results that are easily interpreted and converted to digital format, however, there are disadvantages with DNA sequencing in relation to cost and time issues.

1.9.3.2 Other VNTR loci

In addition to the *spa* repeat region, other genes encoding VNTR loci have been used individually for genotyping, the main gene being the keratin and fibrinogen-binding clumping factor B (*clfB*) (Koreen *et al.*, 2005). Like *spa* typing, sequence analysis revealed that the *clfB* region provides high genotype resolution. The *clfB* repeat region consists of serine-aspartate (SD) repeats and spans between 420 to 804bp. The repeat units are 18bp in length, TCN-GAY-TCN-GAY-AGY-GAY, with N equalling A, C, G, or T and Y equalling C or T. Analysis of a diverse and well characterised collection found that *clfB* typing produced genotypes that correlate with PFGE and microarray analysis. Overall, the method performed well, but was recommended for

use in combination with *spa* typing. This particular combination was tested in another study and was found not to offer any increase in discriminatory power compared with *spa* typing in isolation (Harmsen *et al.*, 2005).

1.9.3.3 Multiple locus VNTR Analysis (MLVA)

Multiple-locus VNTR analysis (MLVA) for *S. aureus* has been developed by two groups who report rapid and high resolution genotyping methods. The first report by Sabat and co-workers used multiplex PCR to amplify the *sdr*, *clfA*, *clfB*, *ssp*, and *spa* loci (Sabat *et al.*, 2003). As the *sdr* loci can have up to three individual VNTR regions, this method produced gel based banding patterns as shown in Fig 1.4. Although the method was tested on a small collection of isolates, it was found to produce equal discriminatory power and reproducibility compared with PFGE. A subsequent report on a larger collection could reproduce the correlation with PFGE and also demonstrated general concordance with the lower resolution methods of MLST and *spa* typing (Malachowa *et al.*, 2005). This study also developed empirical cut-off values for interpretation of the banding patterns that influenced the resulting genotypes. The second *S. aureus* MLVA method, developed by Franscois and co-workers, differs due to automation and inclusion of *fnBP, cna* and *mec*A (Francois *et al.*, 2005). This method benefits from implementation of microcapillary electrophoresis and automatic cluster analysis, facilitating faster experimental turn-around-time. This expanded method further demonstrates the potential of MLVA as a replacement to PFGE, due to automation and high resolution.
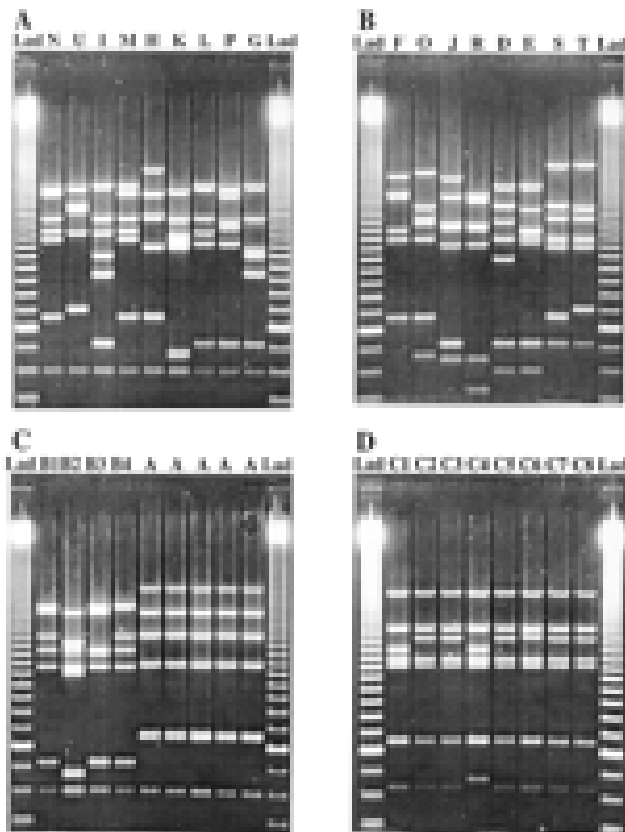
Fig 1.4. MLVA experimental output. The MLVA electrophoretic patterns produce reproducible and high genotyping resolution, however, like PFGE they can be difficult to compare against known genotypes and require gel electrophoresis. Figure from Sabat and co-workers (Sabat *et al.*, 2003).

Independent testing of MLVA for *S. aureus* genotyping was conducted by Tenover and co-workers to determine whether the methodology and results were sufficient to replace PFGE (Tenover *et al.*, 2007). A well-characterised collection of 103 MRSA isolates were tested and revealed that MLVA results were reproducible, but did not differentiate the major United States PFGE lineages to an adequate extent. Although MLVA produced a large number of genotypes, it was not able to separate the CA-MRSA lineages (USA300, 400, 1000 and 1100) from HA-MRSA strains (USA100, 200 and 500). The conclusion for MLVA was that in particular circumstances it could be useful for discriminating among isolates with identical PFGE profiles.

Overall, MLVA is a rapid and cost effective method for generating high resolution genotypes for *S. aureus*. The loci are stable enough to reveal long term relationships, while being sufficiently divergent to subtype major lineages. Automation, data digitisation and high throughput of this method are also advantageous. Disadvantages include inter-laboratory comparison difficulties due to results being gel based banding patterns. Additionally, the current MLVA methodologies neglect sequence variation in individual repeat units as the banding patterns are solely based on repeat region size. Although promising, it remains to be seen whether MLVA becomes conducted routinely for MRSA and MSSA genotyping.

1.9.4 SCCmec Genotyping

The SCC*mec* element is a recently characterised chromosomal island that is responsible for transmission of β-lactam resistance. The SCC*mec* locus varies considerably in content and length and therefore is suited for genotypic analysis. In addition to defining MRSA clones, inherent variation and mobility is invaluable for augmenting the resolution of other methods, such as MLST and PFGE (Enright *et al.*, 2002; Oliveira *et al.*, 2001). Considering the benefits resulting from MRSA genotyping, a variety of methods have been developed for this task.

1.9.4.1 Simplex Method

The original method for characterising the SCC*mec* element is referred to as the "simplex method" and is based around PCR amplification using primers specific for the main *ccr* and *mec* complexes, in addition to the J1 region (Okuma *et al.*, 2002; Ito *et al.*, 2004). The strength of this method is that the five major structural types can be resolved based on their core components. The disadvantages, however, are low resolution resulting from not surveying the extensive observed variation in the junkyard regions, and the number of reactions required. Therefore, alternative methods for SCC*mec* genotyping were developed.

1.9.4.2 Multiplex method

The major alternative to the simplex SCC*mec* genotyping method was the simple multiplex method developed by Oliveira and co-workers (Oliveira and de Lencastre, 2002). In contrast to the simplex method, the SCC*mec* multiplex does not target each of the *ccr* and *mec* classes. Instead nine targets were selected that together classify the SCC*mec* types based upon linkage association with the *ccr/mec* class combinations. The targets include; *mec*A, a region downstream of the *pls* gene, the *kdp* operon, the *mec*I gene, the *dcs* region, a region between pI258 and Tn*554*, a region between Tn*554* and *orfX* and junctions between IS*431* and the plasmids pUB110 and pT181. Numerous studies have employed this method for SCC*mec* characterisation due to standardisation and experimental ease of use. However, the main issue with the multiplex method is the indirect association of the targets with the core structural features that can lead to mistyping of novel *ccr/mec* class combinations such as type V (5C). A further shortcoming of the multiplex method is a lack of resolution for the common type IV (2B) subtypes. Overall this method could performed the role of a rapid and simple method for SCC*mec* genotyping adequately, however, like the simplex method it misses the opportunity to genotype the many known subtypes.

1.9.4.3 Amended Simplex and Multiplex Methods

As an extensive amount of SCC*mec* variation has accumulated, the simplex and multiplex methods have both been upgraded to increase their resolving power. The upgraded simplex methodology features 27 PCRs within six multiplex reactions that incorporate numerous features including junkyard regions (Kondo *et al.*, 2007). However, addition of a single target to produce a single additional genotype has resulted in a laborious undertaking that many clinical microbiology laboratories may find excessive. The multiplex method has been upgraded twice in an attempt to rectify the mistyping issues and to genotype more of the newly discovered variants. The first upgrade was the development of a seven PCR single multiplex for subtyping the type IV (2B) variants (Milheirico *et al.*, 2007a), and the second upgrade featured 10 PCRs that facilitated the differentiation of type V (5C) (Milheirico *et al.*, 2007b). However, like the expanded simplex method, many targets were added that did not add significantly to overall resolution. In an effort to genotype the numerous SCC*mec*

variants, these two methods became overly complicated by inefficiently increasing the number of reactions required. As such, alternative SCC*mec* genotyping techniques have been developed.

1.9.4.4 Other SCCmec Genotyping Methods

One alternate SCC*mec* characterisation methods utilised real-time PCR and multiplexed fluorescent probes to amplify unique sequences from the *ccr*B alleles that correspond with particular SCC*mec* types (Francois *et al.*, 2004). This method has the advantages of being rapid, not relying upon gel electrophoresis and is reasonably priced at US$6-7 per isolate, tested in triplicate. However, the main disadvantage of this method is a low genotyping power, as only SCC*mec* types 1A (I), 2A (II), 3A (III) and 2B (IV) are resolved. A similar method termed the SCC*mec* M-PCR typing assay uses agarose gel based multiplexed PCRs to target unique and specific loci for SCC*mec* types 1A (I), 2A (II), 3A (III), 2B (IV) a, b, c, d, and 5C (V) (Zhang *et al.*, 2005). The method featuring nine pairs of PCR primers was validated on a collection of 453 MRSA isolates and displayed excellent reproducibility. Although performing well, it is limited in its ability to characterise SCC*mec* variants and relies on gel based analysis of PCR banding profiles. An additional gel based multiplex PCR method has been reported and uses only four reactions to define the five major types (Boye *et al.*, 2007). This method was validated on 312 isolates and is simple to perform, however, it does not define SCC*mec* subtypes.

Multiple genotyping methods utilising sequence polymorphisms at the SCC*mec* right extremity junction (MREJ) have also been developed. The method described by Ito and co-workers uses PCR to amplify unique MREJ sequences that correlate with SCC*mec* types 1A (I), 2A (II), 3A (III) (Ito *et al.*, 2001). A more recent method used real-time PCR and molecular beacon probes specific to five MREJ regions, which were found to produce seven genotypes that did not correlate with SCC*mec* types (Huletsky *et al.*, 2004). However, a further study using a single primer pair and DNA sequencing was able to predict the major SCC*mec* structural types and a selection of 2B (IV) subtypes accurately (Cuny and Witte, 2005). Although these methods extract useful information from a single locus, they are not preferred compared with the simplex or multiplex methodologies.

Additional DNA sequenced-based methods for genotyping SCC*mec* have also been developed. One study analysed the *ccr*A and *ccr*B sequences from a diverse collection of MRSA and found that *ccr*B, in particular, was useful for differentiating SCC*mec* types 1A (I), 2A (II), 3A (III), 2B (IV) (Oliveira *et al.*, 2006b). It was an interesting finding that the 2A and 2B *ccr* types could be resolved repeatedly, as they are identical using the simplex *ccr*2 specific PCR. Furthermore, *ccr*B sequences from a number of SCC*mec* 2B variants were identical. This conservation is unsurprising, as the 2B *ccr*2 complex, unlike 2A, is still functional and therefore facilitates SCC*mec* mobility within CA-MRSA strains. Disadvantages of this approach are that SCC*mec* 5C variants are excluded from this analysis as it solely harbours *ccr*5 (*ccr*C), and DNA sequencing is required.

A final method is based on intra-SCC*mec* multi-locus sequencing (Lina *et al.*, 2006). This study indentified four genes that displayed sufficient variation to be used in a sequenced based typing scheme; these include a helicase-like protein, *ccr*A, *ccr*B and a conserved protein of unknown function. The results of the multi-locus sequence typing of SCC*mec* correlated well with the five major SCC*mec* types and produced a total of 21 different SCC*mec* sequence types. The concatenated SCC*mec* sequences were also useful for detecting a low rate of recombination and for determining that no correlation existed with MRSA lineages. SCC*mec* multi-locus sequencing provided a unique insight into the structural evolution of the element and highlighted the use of SCC*mec* genotyping as an efficient method for differentiating MRSA lineages.

As SCC*mec* characterisation is central for investigating MRSA evolution and for tracking problematic strains, many methods have sought to genotype the SCC*mec* element effectively. No current method strikes a balance between resolution, technical efficiency, experimental simplicity and cost. Considering the SCC*mec* variation observed to date, only an excessive number of genotyping targets or the multi-locus sequencing of internal fragments can utilise the diversity to produce high resolution genotype outcomes.

1.10 AIMS OF THIS STUDY

Genotyping surveillance of methicillin-resistant *Staphylococcus aureus* is paramount for managing the spread of hyper virulent and multiresistant strains. MRSA surveillance includes tracking epidemic strains on an intercontinental scale to routine screening of high risk nosocomial departments. MRSA disease is typically associated with health care related patients who are predisposed to opportunistic infection due to immuno-compromisation. Furthermore, as MRSA appears to be replacing MSSA in the general public, genotypic surveillance of *S. aureus* is imperative.

PFGE is the clinical gold standard for MRSA genotyping. Due to issues associated with reproducibility and data output, however, numerous alternate methods have been developed and trialled. MLST, SCC*mec* and *spa* VNTR typing are the modern methods of choice and together produce comprehensive and high resolution genotypes. However, it is a significant undertaking to perform this combination of methods, particularly on larger collections. For this reason, PFGE remains in use. Therefore, the fundamental aim of the current study was to produce a cost effective, high resolution and integrated MRSA genotyping strategy that can be performed routinely using standard laboratory equipment.

The specific aims of this study were:

- To establish an alternate MLST genotyping method based on the interrogation of highly informative SNPs using allele-specific real-time PCR (Chapter Three).
- To systematically identify a small set of SCC*mec* binary targets that maximally genotype the known SCC*mec* variants (Chapter Four).
- To develop a real-time PCR high resolution melting procedure that extracts high genotyping power from the *spa* VNTR without the need for DNA sequencing (Chapter Five).

CHAPTER TWO

GENERAL MATERIALS AND METHODS

2.1 BACTERIAL GROWTH

Bacterial growth for the purpose of DNA extraction was as follows: cryovials containing liquid glycerol stock solutions at -83°C were thawed progressively at -20°C and then at 4°C. Once thawed, 20μL of stock solution was streaked onto a Brain Heart Infusion (BHI) agar plate (prepared as specified, Oxoid Australia Pty Ltd, Adelaide) and cultured overnight at 37°C. The following day a single colony was selected and suspended in a 5mL falcon tube (Becton Dickinson, New South Wales, Australia) containing BHI broth (prepared as specified, Oxoid Australia Pty Ltd, Adelaide) and cultured overnight at 37°C, with shaking.

2.2 DNA PURIFICATION

From the overnight BHI broth culture, 1mL was extracted using the Qiagen DNA extraction kit (Qiagen, Victoria, Australia), as per manufacture's instructions, including lysostaphin at 200μg/mL for the lysis step. Purified DNA samples were eluted using ddH$_2$O and stored at -20°C.

2.2.1 Quantification of DNA

Concentration and purity of DNA samples from the Qiagen DNA extraction kit were determined by measuring the absorbance at 260 and 280nm using an Eppendorf Biophotometer, with each DNA sample diluted 5/100.

2.3 DNA AMPLIFICATION AND ANALYSIS

2.3.1 Polymerase Chain Reaction (PCR)

Standard PCR amplifications were performed using a MJ Research Thermocycler (GeneWorks, Adelaide, Australia) in 0.2mL PCR tubes. The precise reagent and thermocycling specifications are listed in the method sections for Chapters Three, Four and Five. Real-time PCR amplification and detection was conducted on the Applied Biosystems ABI 7000, unless otherwise stated.

2.3.1.1 Primer Design and Synthesis

Primers for both standard and real-time PCR were identified initially using Primer Express (Applied Biosystems, Victoria, Australia), then assessed using the online Net Primer (www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html). During the course of this project, PCR primers were synthesised commercially by Proligo (http://www.proligo.com/) and Geneworks (www.geneworks.com.au/).

2.3.2 Agarose Gel Electrophoresis

Prior to agarose gel electrophoresis, PCR products were combined with loading buffer consisting of 40% glycerol, 0.25% bromophenol blue and 0.25% xylene cyanol (6x). PCR products were then visualised in 1-2% agarose gels that were electrophoresed in TBE buffer (90mM Tris-borate, 2mM EDTA) at 120 volts for 30 - 40min in the presence of ethidium bromide. PCR products were sized against a molecular weight marker (Marker VIII and X, Roche Diagnostics, New South Wales). Approximately 5μL of PCR product was adequate to determine presence and amplification quality. Electrophoresed DNA bands were visualised under ultra-violet light and recorded using a photographic transilluminator (UVP Gel documentation system, USA).

2.4 DNA SEQUENCE ANALYSIS

PCR products were initially prepared for DNA sequence analysis by purification using the QIAquick PCR purification kit (Qiagen, Victoria, Australia). Subsequent quantification was carried out using a 1% agarose gel and comparison against the Roche VIII DNA mass ladder (Roche diagnostics, New South Wales, Australia). The sequencing mix contained an estimated 100ng purified PCR product, 3.2pmoles primer, 4μL ABI Prism BigDye Terminator Mix (version 3.1) and ddH$_2$0 to a final volume of 20μL. The mix was then subjected to sequencing PCR conditions of 94°C for 30s, 50°C for 15s and 60°C for 4min for 25 cycles.

Sequencing products were precipitated in 50μl of 95% ethanol and 2μL of 3M sodium acetate (pH 5.2) at room temperature for 15min, after which the samples were centrifuged at a maximum speed for 20min. The supernatant was carefully removed and 250μl of 70% ethanol added. After centrifugation for 5min at maximum speed, the ethanol was removed and the pellet air-dried. Samples were sequenced at the Australian Genome Research Facility (AGRF, University of Queensland, Brisbane, Australia).

The purified DNA (PD) sequencing method was also used in this study. The PCR products were initially treated with Exo-SapIt (GE Healthcare Bio-Sciences Pty. Ltd., Sydney, Australia) to remove single stranded primer fragments and unincorporated dNTPs. Proceeding quantification, about 2-40ng of purified PCR product and 6.4pmol primer were combined and made up to a volume of 8μL with ddH$_2$0, in a 1.5mL PCR tube. The ABI Prism BigDye Terminator Mix amplification and capillary electrophoresis was conducted by the AGRF.

CHAPTER THREE

GENOTYPING METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* USING MULTI-LOCUS SEQUENCE TYPING SINGLE NUCLEOTIDE POLYMORPHISMS IN COMBINATION WITH BINARY MARKERS

3.1 INTRODUCTION

Methicillin-resistant *Staphylococcus aureus* (MRSA) continues to be a significant human pathogen. For many years it has been the most common cause of nosocomial complications, and the increase in MRSA variants causing infections in the community is an emerging and serious public-health issue (Chambers, 2001; Diekema *et al.*, 2001). Consequently, in recent years there has been an explosion of information on comparative genomics and population biology of *S. aureus*. This has been followed closely by an equally rapid increase in the understanding of the relationship between genotype and clinical phenotype (Robinson and Enright, 2004b; Baba *et al.*, 2002).

Whole genome sequencing and microarray analysis has recently revealed that the *S. aureus* genome includes a backbone of core genes that are essentially common to all variants within the species, together with the accessory genome that consists of genes that are found in some strains but not others (Fitzgerald *et al.*, 2001; Kuroda *et al.*, 2001). Multi-locus sequence typing (MLST) utilises targets within the core component to effectively reveal large scale population structures (Enright *et al.*, 2000; Maiden *et al.*, 1998). MLST studies of diverse *S. aureus* collections have revealed a clonal core genome that evolves primarily via point mutation in concert with rare recombination events (Robinson and Enright, 2004a; Feil *et al.*, 2003). The success of MLST as a *S. aureus* genotyping method is demonstrated by 1057 unique ST sequences at www.mlst.net. However, MLST would appear unsuitable for routine, high throughput use since it requires determination of 3198bp of sequence and does not interrogate the accessory genome.

The fundamental aim of my project was to develop microbial genotyping methods based upon small sets of polymorphisms selected from known genomic diversity on the basis of their optimal combinatorial informative power. A computerised method for identifying such SNPs from MLST databases that provide high Simpson's index of diversity (*D*) values with respect to the databases has been reported (Robertson *et al.*, 2004). Upon application to *S. aureus*, a set of seven SNPs were identified that provide a *D* of 0·95 with respect to the MLST database. To determine the allelic

identity at each polymorphic site, allele-specific real-time PCR assays were reduced to practice and tested on a limited number of isolates. However, Robertson and co-workers did not determine the concordance between SNP profiles and *S. aureus* population structure, i.e. whether the SNP-based genotyping clusters closely related isolates together and discriminates among more distantly related ones. Accordingly, the first aim of this study was to address this question as it relates to MRSA and to assess the performance of the SNP-based genotyping method on a larger collection of MRSA isolates.

Clinical laboratories ideally require molecular genotyping methods that are convenient, economical and accurately resolve the position of the isolate within the population structure. The 'progressive hierarchical resolving assays using nucleic acids' (PHRANA) concept articulated by Keim and co-workers is very helpful in devising such typing methods (Keim *et al.*, 2004). The underlying principle of PHRANA is that slowly evolving polymorphic sites do not provide high resolution but are effective at revealing the location of an isolate within the population structure of the species as a whole. While rapidly evolving markers provide high resolution but are unreliable at indicating the position of the isolate within the population structure, primarily because of homoplasy (identical allelic states arising through independent routes). However, the use of rapidly and slowly evolving markers in combination provides a usefully complete picture. Therefore, additional aims in this chapter were to capitalise on the broad, but low resolution high *D* SNP set by adding targets from loci that evolve at faster rates. It was hypothesised that choosing binary targets associated with virulence and antimicrobial resistance would also provide inferences regarding clinical phenotype.

It was envisaged that a *S. aureus* genotyping method based on the interrogation of a set of MLST database-derived SNPs in combination with a set of variable virulence genes would provide efficient and informative genotyping resolution. The primary guide to the selection of virulence factor encoding genes was a study by Peacock and co-workers who identified a number of such genes exhibiting binary variation within MLST clonal complexes (Peacock *et al.*, 2002).

In addition to elevated virulence, the most clinically hazardous *S. aureus* are those expressing antimicrobial resistance, in particular the strains harbouring SCC*mec* (Ito *et al.*, 2001). Numerous studies based on characterising SCC*mec* variants have revealed extensive diversity. Accordingly, SCC*mec* binary targets are commonly used in combination with MLST to define the MRSA clones (Oliveira *et al.*, 2001). In considering alternative genotyping targets for combinatorial use with the SNP set, it was hypothesised that SCC*mec* binary targets would be potentially suitable. Therefore, a further aim was to characterise the SCC*mec* elements in our collection and compare the informative powers of the virulence genes and the SCC*mec* targets, in combination with the genotyping SNPs.

3.2 MATERIALS AND METHODS

3.2.1 Bacterial Isolates

One hundred and seven MRSA isolates from four collections were included in this study. Forty-seven isolates were from the Australian Group on Antimicrobial Resistance (AGAR) 2000 community-onset collection (Coombs *et al.*, 2004), 15 isolates were from a cluster of community-onset disease in South-East Queensland, Australia, nine were from another similar cluster of community-onset disease, and 36 isolates have been described previously (Huygens *et al.*, 2004; Nimmo *et al.*, 2000). Complete descriptions of the isolates are available in the Appendix Table 1.

3.2.2 High D SNP Genotyping Set

A highly discriminatory set of seven SNPs derived from the *S. aureus* MLST database has been reported previously (Robertson *et al.*, 2004). Identification of the SNPs was facilitated by the computer program Minimum SNPs, which can take as input a complete MLST database and provide as output sets of key genotyping SNPs that, if interrogated, provide a high Simpson's index of diversity, *D*. We term this a 'high *D*' SNP set. The bioinformatic process of identifying the *S. aureus* high *D* SNP set using Minimum SNPs has been reported previously (Robertson *et al.*, 2004).

3.2.3 Interrogation of the High D SNP Set using Allele-Specific Real-Time PCR

All allele-specific real-time PCR reactions were carried out in an Applied Biosystems ABI 7000 Sequence Detection system using Applied Biosystems SYBR Green PCR MasterMix (Applied Biosystems, Victoria, Australia). Unlabelled primers listed in Table 3.1 were obtained from Proligo (http://www.proligo.com/) and purified by desalting only (Robertson *et al.*, 2004).

Table 3.1. Primer sequences for the *S. aureus* high *D* SNP set.

SNP	Primer*	5` Sequence 3`
arcC210	arcC210-C	CGTATAAAAAGGACCAATTGGTTT<u>G</u>
	arcC210-T	CGTATAAAAAGGACCAATTGGTTT<u>A</u>
	arcC210-A	CGTATAAAAAGGACCAATTGGTTT<u>T</u>
	arcC210-For	TATGATAGGCTATTGGTTGGAAACTG
tpi243	tpi243-A	GTAAATCATCAACATCTGAAGATGC<u>A</u>
	tpi243-G	GTAAATCATCAACATCTGAAGATGC<u>G</u>
	tpi243-Rev	CTTCTTTGCTTGATAAGTCAGCAATAG
arcC162	arcC162-T	GTGATAGAACTGTAGGCACAATCGT<u>T</u>
	arcC162-A	GTGATAGAACTGTAGGCACAATCGT<u>A</u>
	arcC162-Rev	GGGTTATTGAATCGTGGATCATC
tpi241	tpi241-G	GGTAAATCATCAACATCTGAAGAT<u>G</u>
	tpi241-A	GGTAAATCATCAACATCTGAAGAT<u>A</u>
	tpi241-Rev	CTTCTTTGCTTGATAAGTCAGCAATAG
yqiL333	yqiL333-C	TGCTTGTCAACAACAGTCGCTT<u>C</u>
	yqiL333-T	TGCTTGTCAACAACAGTCGCTT<u>T</u>
	yqiL333-Rev	TCTGTTAAACCATCATATACCATGCTATC
aroE132	aroE132-A	GGCTTTAATATCACAATTCCTCATAAAGA<u>A</u>
	aroE132-G	GGCTTTAATATCACAATTCCTCATAAAGA<u>G</u>
	aroE132-Rev	CTTGTCATCTTTTATCAAAACAGTGTTAAC
gmk129	gmk129-C	GGATGCGTTTGAAGCTTTAAT<u>C</u>
	gmk129-T	GGATGCGTTTGAAGCTTTAAT<u>T</u>
	gmk129-Rev	TTGTATCTTTAACATATTGAACTGGTGTAC

*Allele-specific primers are named according to the position of the SNP and the base in the sense strand at the SNP. The 3' end of these primers always coincides with the SNP. Common primers are named according to the position of the SNP and whether the primer is oriented for the antisense strand (-Rev) or the sense strand (-For).

Each 20μl allele-specific real-time PCR reaction was conducted in duplicate and contained; 10μl 2x SYBR green PCR master mix (Applied Biosystems, Victoria, Australia), 0.2μM of each forward and reverse primer, and 6.8μl DNA template containing approximately 10ng of purified template DNA. Each SNP was amplified using a three-step temperature cycling procedure as follows: 50°C for 2min and 95°C for 10min, followed by 40 cycles of 95°C for 15s, 52°C for 20s and 72°C for 33s. The integrity of the allele-specific PCRs were checked routinely by ensuring that the amplicon melting-points were consistent with the size of the expected product and not with primer-dimers, and also to ensure that a single product was produced.

The Δ cycle time (ΔC_T) values were calculated by subtracting the CT of the matched primer/template pair from the CT of the mismatched primer template. CT values were obtained by manually setting the fluorescence threshold to the exponential amplification phase. The premise of allele-specific real-time PCR is shown in Fig 3.1.
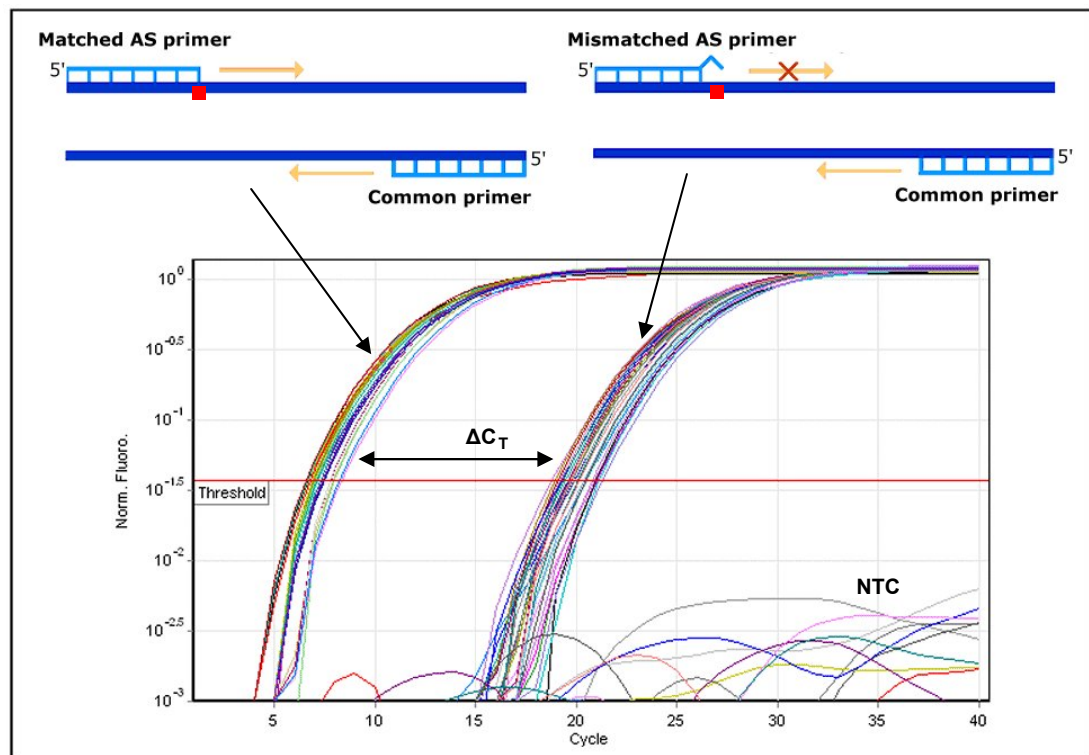
Fig 3.1. SNP identification using allele-specific real-time PCR. For each potential nucleotide at the SNP position (red dot), a matched primer with a corresponding 3` nucleotide is required. When each primer is tested individually, only the matched reaction will amplify efficiently as *Taq* DNA polymerase is hindered by uncomplimentary 3` nucleotides (mismatched primers), thus producing inefficient amplification. The allele-specific reaction that first reaches the fluorescent threshold determines the SNP identity. The amplification efficiency between reactions is measured quantitatively by the difference in cycles to threshold (ΔC_T). NTC = no template controls. Figure adapted from E. P Price with permission.

3.2.4 Binary Marker Selection, Primer Design and PCR Amplification

Candidate virulence genes for instigating SNP clonal complex subtyping were selected based on previous studies (Moore and Lindsay, 2001; Peacock *et al.*, 2002). Real-time PCR primers for the genes *fnbA*, *cna*, *icaA*, *sdrE*, *bbp*, *hlg*, *pvl* and *tst* were designed using the Primer Express 2.0 software package (Applied Biosystems, Victoria, Australia) (Table 3.2). Virulence gene detection was conducted on an ABI7000 real-time PCR device with the following PCR cycling conditions: 50°C for 2min, 95°C for 10min; then 30 cycles of 95°C for 15sec, and 60°C for 1min. Each 20µl reaction was conducted in duplicate and contained 10µl 2x SYBR green PCR master mix (Applied Biosystems, Victoria, Australia), 0.125µM each forward and reverse primer, and 8µl DNA template containing approximately 10ng of purified template DNA. Primer sequences and PCR conditions for amplifying the plasmids pT181, pUB110 and pI258 have previously been published (Huygens *et al.*, 2002).

Table 3.2. Primer sequences for the binary virulence genes.

Virulence Gene	Gene function	GenBank accession no.	Primer	5` Sequence 3`
fnbA	Fibronectin binding protein	J04151	fnbA_F1	ACA GTA ACA GAA CAA CCG TCA AAC G
			fnbA_R1	TTG CTG GTT GTG CAG TTT GTG
cna	Collagen adhesin protein	M81736	cna_F1	TCA AGC AGT TAT TAC ACC AGA CGG
			cna_R1	GCA AAT CCC GAA ACA TCA CTT AA
sdrE	Putative adhesin	AJ005647	sdrE_F2	CAT CAG ATG GCG AAC TAG ATA TTG C
			sdrE_R2	CCA GTG TCA TTA GAA GTT ACG ATG AAG
bbp	Bone binding protein	Y18653	bbp_F1	AAT GAC AGT AAA GTA TGA CAA GGA GAC G
			bbp_R1	TTC CCT TCG CTA TTA CTT CAC CTG
icaA	Intracellular adhesin	AF086783	icaA_F1	TCG ACG TTG GCT ACT GGG AT
			icaA_R1	GGC AAG CGG TTC ATA CTT AAT ACG
pvl	Leukocidin	X72700	*pvl*_F1	TCA CAA AAT GCC AGT GTT ATC CAG
			*pvl*_R1	TTT TGC AGC GTT TTG TTT TCG
hlg	Hemolysin	L01055	hlg_F1	ATG AAA TCA AGG TGA AAG GAC AGA A
			hlg_R1	GAC TTT TTT TAC GCT GAC TGG TGT
tst	Toxic shock syndrome toxin	AY074881	tst_F1	CCC TTT GTT GCT TGC GAC A
			tst_R1	CCC ACT ACT ATA CCA GTC TAG CAA ATC C

3.2.5   Multi-locus Sequence Typing (MLST)

MLST PCRs were conducted as described previously (Enright *et al.*, 2000). Sequencing reactions were carried out using ABI PRISM BigDye Terminator mix (version 3.1) according to the manufacturer's instructions. Sequencing products were electrophoresed at the AGRF. Twenty-four isolates from the collection used in this study have been subjected to MLST genotyping previously (Coombs *et al.*, 2004).

3.2.6   eBURST Population Structure Analysis

The MLST population structure for determining the resolving power of the high *D* SNP set was constructed using eBURST as specified by Feil and co-workers (Feil *et al.*, 2004). Briefly, to reconstruct clonal complexes based on MLST sequence data, the eBURST algorithm first identifies clonal groups of STs that share similar allelic profiles. The clonal progenitor is then identified as the ST displaying the maximum number of SLVs. This progenitor ST is then displayed diagrammatically in the centre of the clonal complex and the SLVs and DLVs are displayed one and two stops from the progenitor ST, respectively (Fig 3.2).
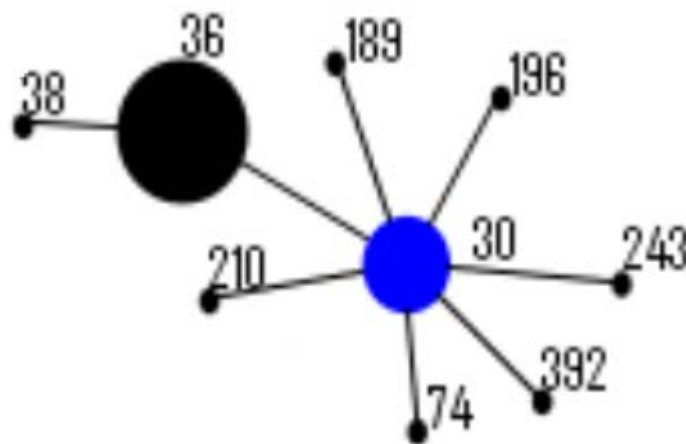
Fig 3.2. The eBURST output for MLST clonal complexes. In this example, the ST30 is the clonal complex progenitor, which incurred a single locus change to produce the successful ST36 (SLV) that in turn accrued another mutation producing ST38 (DLV). The eBURST depiction of ST relatedness does not take into account the overall number or type of SNPs, only the total allele differences. The size of the dots representing each ST is proportional to the number of STs of that type in the MLST data set.

3.2.7   SCCmec Genotyping

The SCC*mec* structural types were determined by PCR identification of the *ccr* allele, the *mec* complex class and in some cases the J1 region, as described previously (Okuma *et al.*, 2002; Ito *et al.*, 2004). Twenty-four isolates characterised previously for SCC*mec* are described elsewhere (Coombs *et al.*, 2004).

3.2.8   agr Genotyping

Four different *agr* types were identified using the multiplex PCR procedure described previously (Lina *et al.*, 2003).

3.3 RESULTS

3.3.1 Informative Power of the Seven-member High D SNP Set for MRSA Genotyping

It is desirable that any genotyping method provides outputs that are consistent with the population structure of the species in question (i.e. the similarity of genotypes is correlated with the relationship of isolates), as this facilitates the inference of dissemination and evolutionary events. Accordingly, the consistency between the MRSA population structure and the SNP profiles was first determined by Dr. Flavia Huygens (Fig. 3.3). The population structure is depicted using eBURST analysis of MRSA sequence types present in the *S. aureus* MLST database as at January 25, 2005. eBURST was used because conventional phylogenetic analyses of MLST databases may be confounded by recombination events. The eBURST algorithm identifies clonal complexes using MLST allelic data and attempts to identify clonal complex progenitors (Feil *et al.*, 2004). This analysis showed that seven SNPs resolve known human-derived MRSA STs into 29 SNP profiles. Major HA- and CA-MRSA clones are discriminated, with the exception of the unrelated STs 59 and 93, which represent the North American SF25 : P (Diep *et al.*, 2004) and Queensland clones (Munckhof *et al.*, 2003) of CA-MRSA, respectively. Even though the complete set of STs includes many that were not known at the time that the SNP set was defined, the SNP genotypes are concordant with the known population structure.

Fig 3.3. Output of an eBURST analysis of *S. aureus* STs in the MLST database listed as being MRSA of human origin. The group definition was set to 'zero alleles in common' in order to allow visualization of all STs on a single diagram. Solid lines represent single locus variation, and double locus variation is not indicated. Each circle represents a particular SNP profile. The SNP profiles are indicated in the order *arcC*210, *tpi*243, *arcC*162, *tpi*241, *yqiL*333, *aroE*132, *gmk*129. The diameters of the solid spots are proportional to the occurrences of the STs in the MLST database, i.e. a large spot means that a large number of isolates with that ST have been entered into the database. Actual SNP profiles found in this study are boxed, together with the number of isolates.

3.3.2   Genotyping Known Australian MRSA using the High D SNP Set

To investigate the resolving power of the high *D* SNP set with actual isolates as well as the robustness of the kinetic PCR method for interrogating these SNPs, a selection of Australian MRSA isolates were subject to SNP genotyping. This experimental work was completed in conjunction with Dr. Flavia Huygens. Initially, the SNP interrogation method was validated using 32 MRSA isolates of known ST (see Appendix Table 1). The kinetic PCR was completely successful at calling the bases at each SNP. For every polymorph (allelic state of a SNP), the reaction with the perfectly matched primer reached the threshold in fewer cycles than the reaction with the mismatched primer, i.e. the ΔC_T was always in the expected orientation (Fig 3.4). Based on these trials it was concluded that the kinetic PCR approach to interrogating these SNPs is robust. The 11 STs (1, 8, 22, 30, 45, 73, 78, 88, 93, 128 and 239) were resolved into nine SNP profiles. Unresolved STs (ST239–ST128 and ST78–ST88) are very closely related to each other (Fig. 3.3).
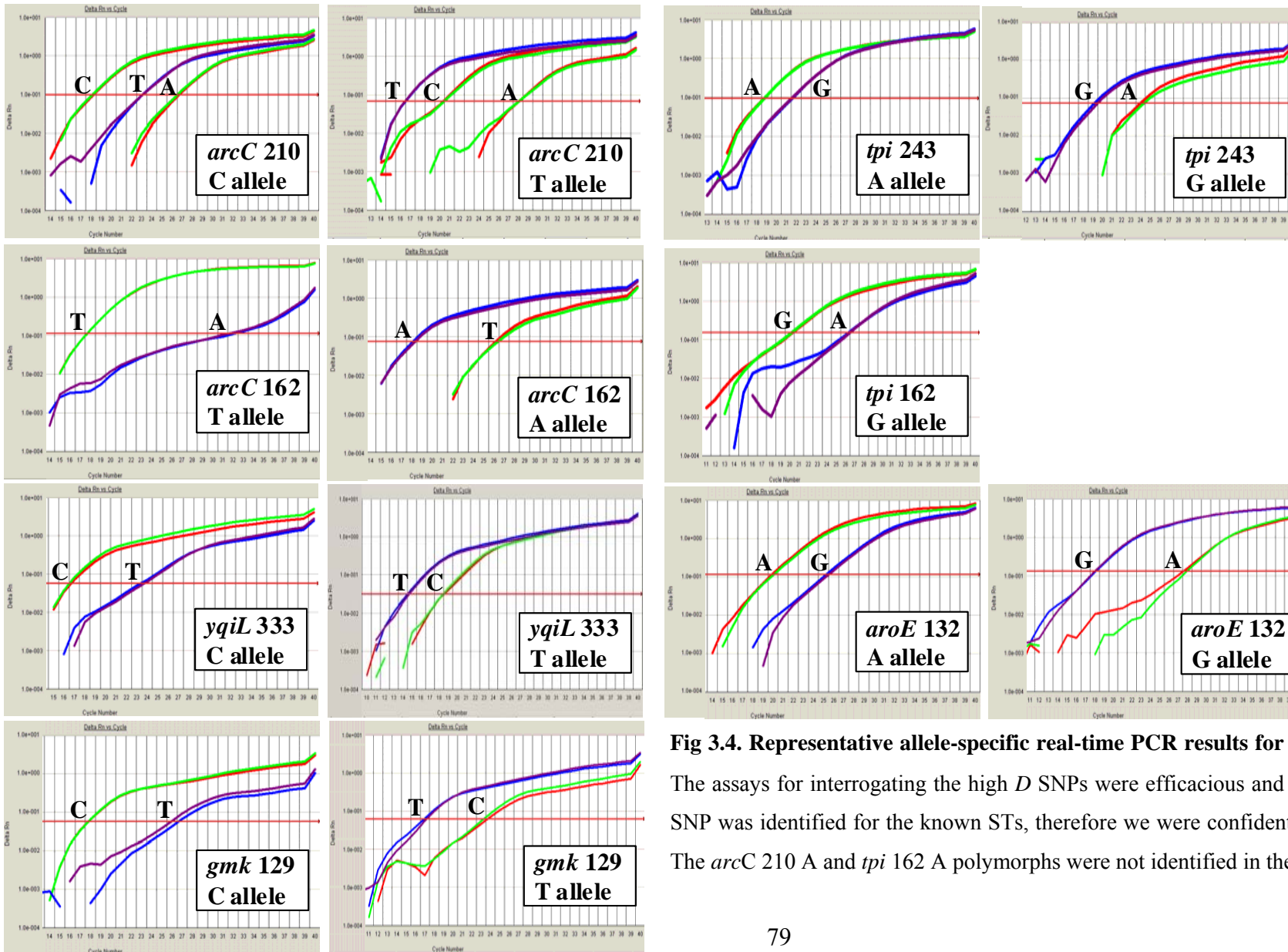
Fig 3.4. Representative allele-specific real-time PCR results for each high *D* SNP polymorph. The assays for interrogating the high *D* SNPs were efficacious and robust. In each case the correct SNP was identified for the known STs, therefore we were confident to genotype unknown isolates. The *arc*C 210 A and *tpi* 162 A polymorphs were not identified in the MRSA collection.

3.3.3 SNP-based Genotyping of Unknown Australian MRSA

The SNP genotyping method was then carried out on 76 MRSA isolates from South-East Queensland and elsewhere in Australia that had not been subject to full ST determination (Table 3.3). Nine SNP profiles were obtained, and these were consistent with ST1, ST73 (ST5), ST8, ST22, ST30, ST45, ST78–ST88, ST93 and ST239 (ST73 was found in the fully sequence-typed isolates, and is a SLV of ST5 that is a very well known, and common, clonal complex progenitor). The SNP profiles showed consistency with the PFGE profiles and origins of the isolates. The ST30/PFGE type A isolates are members of the Western Samoan Phage Pattern (WSPP)/South West Pacific (SWP) clone that is strongly associated with community-acquired infections in Oceania (Nimmo *et al.*, 2000). As expected, PFGE provided higher resolution than the high *D* SNP set, and on no occasion were different SNP profiles associated with the same PFGE type (Table 3.3). As the STs included in this study encompass much of the diversity of Australian MRSA (Coombs *et al.*, 2004), it was concluded that the high *D* SNP set was useful for identifying the major Australian MRSA clones.

Table 3.3. Representative SNP profiles for isolates of unknown ST.

Isolate	ST*	PFGE type	*arcC*210 C/T/A	*tpi*243 A/G	*arcC*162 T/A	*tpi*241 G/A	*yqiL*333 C/T	*aroE*132 A/G	*gmk*129 C/T
MD828534	(1)	E	C	A	T	G	C	A	C
MB827549	(1)	E	C	A	T	G	C	A	C
MD808118	(1)	L	C	A	T	G	C	A	C
E804531	(5)	I	C	A	T	G	C	A	T
K704540	(239)	F	T	A	A	G	T	A	C
K705613	(239)	F2	T	A	A	G	T	A	C
K711532	(239)	F3	T	A	A	G	T	A	C
K714372	(239)	F4	T	A	A	G	T	A	C
MIC703484	(239)	G1	T	A	A	G	T	A	C
I823541	(239)	G2	T	A	A	G	T	A	C
E812560	(239)	J	T	A	A	G	T	A	C
IPOOM14235	(239)	O	T	A	A	G	T	A	C
D828570	(30)	A0	T	G	A	G	T	G	T
MA823547	(30)	A1	T	G	A	G	T	G	T
MC810534	(30)	A1	T	G	A	G	T	G	T
MD821552	(30)	A2	T	G	A	G	T	G	T
E802537	(30)	A3	T	G	A	G	T	G	T
68284/98	(30)	A5	T	G	A	G	T	G	T
E822485	(30)	B	T	G	A	G	T	G	T
MJ710566	(45)	C	C	G	A	G	T	G	T
B8-31	(45)	K	C	G	A	G	T	G	T
16077-5584	(93)	R	T	G	T	G	T	A	T
IP01M2046	(78,88)	P1	T	A	T	G	C	A	T

*The bracketed figures in the column labeled "ST" are the STs found in the isolates of known ST that are consistent with the SNP profile.

3.3.4   Identification of Binary Markers Exhibiting Useful Variability

The high *D* SNP set is effective at deciphering the stable genome backbone and for characterising isolates into MLST clonal complexes. In order to increase resolution and potentially to reveal virulence properties, a small set of genes exhibiting complementary binary variability were identified, i.e. their presence or absence is not predictable either from the SNP profile or from the presence or absence of other binary markers. Ideally, the markers would also have the potential to directly provide clinically significant information. Peacock and co-workers identified a set of *S. aureus* virulence genes with a distribution that did not correlate with ST, and therefore were considered as ideal candidates (Peacock *et al.*, 2002). These genes included *pvl*, *cna*, the *sdrE/bbp* alleles, *tst*, *icaA*, *fnbA* and *hlg* (Table 3.4).

To identify gene variation within the 107 MRSA isolates, real-time PCR assays were designed and optimised. The PCR amplicon length for each reaction was verified using agarose gel electrophoresis. Results from the PCR assays indicate that the *pvl*, *cna*, *sdrE/bbp* and *tst* genes were variable within the collection, and that the *icaA*, *fnbA* and *hlg* genes were present in each isolate (Table 3.4).

Table 3.4. Gene prevalence within *S. aureus* collection.

Gene	Function	Isolate acquisition		Total prevalence
		Community (n=66) (%)	Nosocomial (n=41) (%)	
fnbA	Fibronectin binding protein	66 (100)	41 (100)	107 (100)
cna	Collagen adhesion protein	54 (81.8)	35 (85.4)	89 (83.2)
sdrE	Binding protein	35 (53)	39 (95.1)	74 (69.2)
bbp	Bone binding protein	29 (43.9)	1 (2.4)	30 (28)
icaA	Polysaccharide intercellular adhesin	66 (100)	41 (100)	107 (100)
pvl	Leokocydin	48 (72.7)	0 (0)	48 (44.9)
hlg	Heamolysin	66 (100)	41 (100)	107 (100)
tst	Superantigen	0 (0)	4 (9.8)	4 (3.7)

3.3.5 Binary markers in Combination with High D SNPs Increase Overall Resolution

Given that the *pvl*, *cna*, *sdrE/bbp* and *tst* genes exhibited variation within the collection, it was then necessary to examine their distribution within isolates of identical SNP profiles in order to determine whether subtyping was achievable (Fig 3.5). In regard to the CA-MRSA SNP profiles of ST1, ST30 and ST93, the binary virulence genes produced a total of three, two and two subtypes, respectively. The *pvl* gene is located within transmissible bacteriophages and is therefore known to display variation within closely related MRSA strains. The epidemic (EMRSA16) ST36 is a SLV of ST30 and is not associated with community acquisition. The single ST36 isolate in our collection was differentiated from the CA-MRSA ST30 strain because it was *pvl* negative and *tst* positive. The ST93 SNP profile encompassed two PFGE pulsotypes, Q and R. The ST93 R pulsotype is known as the CA-MRSA Queensland clone and is characteristically *pvl* positive. The pulsotype Q isolate was *pvl* negative and full MLST determination revealed it to be an ST59. Analysis using eBURST indicates that ST59 and ST93 are not closely related, however, they coincidentally share the same SNP profile. This is not fortuitous, as the emerging CA-MRSA SF25 : P clone in the United States is ST59 (Diep *et al.*, 2004). Although the ST59 isolate in this study was *pvl* negative and could be discriminated from ST93 isolates on this

basis, ST59 can also be *pvl* positive (Diep *et al.*, 2004). Thus, it could be difficult to discriminate from the Queensland clone. Because ST93 is a singleton, however, we were able to use Minimum SNPs to determine that ST93 can be discriminated from all other STs in the database by presence of a G at *aroE*252. This SNP could easily be added to a standardised procedure based on the high *D* SNP set, or alternatively could underpin a separate single-SNP assay that is only carried out when the high *D* SNP profile TGTGTAT is obtained.

With respect to nosocomial isolates, the ST239 group consisted of 24 isolates, but was only divided into two subtypes based on virulence gene variation; a single isolate tested negative to both *sdrE* and *bbp*. The two isolates genotyped as belonging to the ST5 SNP profile could be differentiated via the *cna* locus since the ST5 isolate was negative and the ST73 isolate was positive. Binary variation of the *cna* gene was also found to subtype the ST8 and ST78 SNP groups. Variation of the virulence gene targets within the SNP groups of ST22 and 45 was not identified; this however, is most likely due to the small sample sizes examined in these groups.

Overall, the virulence genes did not display the level of variation reported in other studies (Peacock *et al.*, 2002). Nonetheless, a selection of genes, notably *pvl*, displayed utility as potential targets for subtyping SNP groups.

3.3.6   SCCmec Associated Plasmids add Further Resolving Power

It is well known that one mode of SCC*mec* variation involves integrated plasmids within the J3 region (Oliveira *et al.*, 2000). Studies have demonstrated previously that these plasmids may be used for binary typing (Huygens *et al.*, 2002). The resolving power of integrated plasmids in combination with the high *D* SNP set and binary virulence genes were assessed in the current study (Fig 3.5).

Variant 93a – *pvl* positive
93b – *pvl* negative

n = 13

Variant 30a – WSPP
30b – *pvl* negative
tst & pUB110 positive

n = 4

Variant 78a – *cna* positive
78b – *cna* negative

n = 3

Variant 239a – pI258 positive
239b – pI258 + pT181 positive
239c – pUB110 positive
239d –no mec plasmids
239e – *sdrE* and *bbp* negative

n = 30

Variant 1a – *pvl* positive
1b – *pvl* negative
1c – *sdrE* negative
1d – pT181 positive

n = 25

n = 2

Variant 8a – *cna* positive
8b – *cna* negative

n = 3

Variant 5a – *cna* positive
5b – *cna* negative

n = 24

Fig 3.5. eBURST diagram displaying the increase in genotyping resolution provided by the binary virulence and SCC*mec* associated plasmids. The red dots indicate the isolates within the collection that were genotyped using MLST. The STs located within the defined circles indicate the resolution of the high *D* SNP genotyping set. The blue boxes display the number and nature of the subtypes produced from testing *pvl*, *cna*, *sdrE/bbp*, *tst*, pUB110, pT181 and pI258 in combination with the high *D* SNPs.

The entire collection was tested for pUB110, pT181 and pI258 and variation was identified in the ST1, ST30 and ST239 SNP profiles. An additional subtype was identified from the CA-MRSA-associated ST1 SNP profile, with five *pvl*-negative isolates testing positive for pT181. The ST36 isolate sharing the ST30 SNP profile was differentiated further from the ST30 strain because it tested positive for pUB110. Most variation present was in the ST239 SNP profile isolates, among which, four plasmid profiles were found. The majority carried pT181 and pI258, with the remaining isolates carrying either pI258 or pUB110, if they carried a plasmid. It was concluded that these plasmids can increase the efficiency of genotyping resolution, particularly with nosocomial isolates that harbour the large SCC*mec* type III (3A/*mer*) elements.

The overall increase in resolution resulting from targeting binary virulence genes and SCC*mec* plasmids is evident in Fig 3.5.

3.3.7   Comparison of Binary Markers with SCCmec Genotyping

A well accepted genotyping approach uses SCC*mec* typing and MLST determination in combination. SCC*mec* genotyping is a binary system based primarily on variation in the *ccr* gene complex and the *mec* complex (Okuma *et al.*, 2002). It was therefore of interest to compare the informative power of SCC*mec* typing with that of other binary markers.

SCC*mec* characterisation was undertaken on the remaining 83 isolates that had not been genotyped previously. Each of the five major structural types were identified

(SCC*mec* I–V) successfully, along with at least four novel variants. The five ST1 SNP profile isolates that tested negative for *pvl* and positive for pT181 were found to carry a novel SCC*mec* element composed of *ccr*1 with *mec* class C. Two unrelated isolates were found to carry SCC*mec* elements that were positive for *mec* class B and negative for each *ccr* allele. Interestingly, 10 of the 29 ST30 SCC*mec* type IV isolates tested negative for subtype a, as did the ST22 SNP group. The J1 regions of these isolates were analysed further using PCR, but proved negative for subtypes b and c. The ST239 SNP profile isolates tested in this study were each found to carry SCC*mec* type III, but as stated earlier, these elements vary with respect to integrated plasmids that define the subgroups. There was also a single ST1 SNP isolate (FH53) that carried SCC*mec* type I. In conclusion, SNP genotyping used in combination with SCC*mec* types, including the novel variants, defined 17 genotypes. This is fewer than the 21 genotypes defined by the SNPs+binary virulence genes+integrated plasmids, however 24 genotypes were defined when all loci were considered.

3.3.8   Does agr Variation Increase Genotyping Resolution of the High D SNP Set?

The *agr* locus is a polymorphic region of the genome that modulates expression of a set of virulence-related genes (Ji *et al.*, 1995). To determine if variation at this locus is useful for increasing resolution, the collection was assayed for each of the four *agr* types in collaboration with Mr. John Inman-Bamber. A single instance of intra-SNP profile variation at the *agr* locus was identified. The *pvl*-negative ST59 isolate that shares a SNP profile with the ST93 SNP genotype was found to have *agr* type I, compared with the ST93 isolates that carry *agr* type III. These results indicate that with this collection of isolates, binary variation is concordant with genotypes defined by the high *D* SNPs and other binary markers, so interrogation of the *agr* locus adds little additional informative power in this context.

A summary of the genotyping performance of the SNPs, binary targets, SCC*mec* and *agr* typing are collated in Table 3.5.

Table 3.5. High *D* SNP and binary marker combinatorial genotypes obtained in this study.

SNP genotype	No. of isolates	Profile									Clone	Reference
		pvl	*cna*	*sdrE/bbp*	*tst*	pT181	pI258	pUB110	SCC*mec*	*agr*		
CATGCAC												
1a	3	+	+	*sdrE*	−	−	−	−	IV	3	WAMRSA-1	Coombs *et al.* (2004)
1b	11	−	+	*sdrE*	−	−	−	−	IV, I	3	WAMRSA-1	Coombs *et al.* (2004)
1c	5	−	+	*sdrE*	−	+	−	−	New	3		
1d	2	−	+	Neither	−	−	−	−	IV	3		
CATGCAT												
5a	1	−	−	*sdrE*	−	−	−	−	IV	2		
5b	1	−	+	*sdrE*	−	−	−	−	IV	2		
TATGTAC												
8a	2	−	−	*sdrE*	−	−	−	−	IV, V	1	nmMRSAD	Coombs *et al.* (2004)
8b	1	−	+	*sdrE*	−	−	−	−	New	1	Irish-2	Coombs *et al.* (2004)
CGTGTGC												
22	4	−	+	*sdrE*	−	−	−	−	IV	1	EMRSA-15	Coombs *et al.* (2004)
TGAGTGT												
30a	29	+	+	*bbp*	−	−	−	−	IV	3	SWP/WSPP	Coombs *et al.* (2004)
30b	1	−	+	*bbp*	+	−	−	+	II	3	EMRSA-16	Nimmo *et al.* (2000)
CGAGTGT												
45	3	−	+	*sdrE*	+	−	−	−	V	1	nmMRSAE	Coombs *et al.* (2004)
TATGCAT												
78a	2	−	−	*sdrE*	−	−	−	−	IV, New	3		
78b	1	−	+	*sdrE*	−	−	−	−	IV	3		
TGTGTAT												
93a	16	+	−	*sdrE*	−	−	−	−	IV	3	QLD	Coombs *et al.* (2004)
93b	1	−	−	*sdrE*	−	−	−	−	IV	1		
TAAGTAC												
239a	10	−	+	*sdrE*	−	+	+	−	III	1	AUS-3	Coombs *et al.* (2004)
239b	8	−	+	*sdrE*	−	−	−	+	III	1	AUS-2	Coombs *et al.* (2004)
239c	4	−	+	*sdrE*	−	−	+	−	III	1		
239d	1	−	+	*sdrE*	−	−	−	−	III	1		
239e	1	−	+	Neither	−	+	+	−	III	1	AUS-3	Coombs *et al.* (2004)

The SNP profiles are shown in column 1, in the order *arc*C210, *tpi*243, *arc*C162, *tpi*241, *yqiL*333, *aroE*132, *gmk*129. The "clone" column refers to the common names for these lineages based on PFGE.

3.4 DISCUSSION

Prior to the current study, Robertson and co-workers developed a computerised method for identifying small sets of SNPs from MLST databases that resolve lineages or hyper-virulent clones efficiently within bacterial species (Robertson *et al.*, 2004). This SNP identification process using Minimum SNPs software involves concatenating the allelic sequences for each ST in the MLST database, then compiling a mega alignment of the concatenated sequences where SNP sets displaying high *D* values are identified. For *S. aureus*, Robertson and colleagues identified a set of seven SNPs with a *D* value of 0.95 that in effect, is the probability that any two randomly chosen STs from the input data set will be differentiated. The first issue in this study was to determine whether use of the high *D* SNP set preserved evolutionary meaningful relationships within the natural population structure. This was achieved by creating an eBURST-derived population structure consisting of STs and clonal complexes. Although the STs that grouped within SNP profiles correlated well with the factual groups of STs constituting the MLST defined CCs, some anomalies were also apparent.

It is obvious that a small SNP set cannot possess the discriminatory power of full ST determination since the majority of strains are defined by unique SNPs. In Fig. 3.3, it can be seen that there are two related, but different phenomena. Firstly, there are instances in which the SNPs fail to discriminate unrelated singletons or clonal complexes. Increasing the size of the SNP set from seven to 14 confers the ability to resolve the great majority of unrelated clonal complexes and singletons (Dr. Flavia Huygens data not shown). It is reasonable to suggest then, that the optimum size of a SNP set is a function of the need to discriminate singletons and small clonal complexes.

Secondly, the SNP set provided very little power to discriminate among members of clonal complexes. This is to be expected, since *S. aureus* is a clonal organism with low diversity, and the great majority of SLVs of clonal complex progenitors have arisen by mutation (Feil *et al.*, 2003). For a given SLV, the SNP that discriminated it from the clonal complex progenitor is not polymorphic anywhere else in the species,

and so will not be readily identified in a computerised search for high D SNP sets. A SNP-based method for resolving within *S. aureus* clonal complexes as defined by MLST would need to resolve each clonal complex member individually from the clonal complex progenitor, and this would be very inefficient. Consistent with this observation, in the majority of the small number of instances in which the SNP profile of an SLV differed from the SNP profile of the clonal complex progenitor, the SLV had arisen by recombination rather than mutation. For example, the alleles that discriminate STs 221, 194 and 148 from ST5 are found elsewhere in the MLST database, while the corresponding alleles for the other SLVs of ST5 are unique. ST221 is particularly interesting because it has acquired an allele that is characteristic of ST1, and shares four of its seven alleles, as well as a SNP profile, with ST1. In this instance, analysis of the SNP profiles revealed a recombination event.

In addition to *in silico* population analysis, the novel SNP genotyping method required its performance to be assessed *in vitro*. For this reason a diverse collection of known and unknown Australian MRSA isolates were subjected to SNP typing to confirm the robustness of both the allele-specific real-time PCR assays and the resolution of the high D SNP set. From a technical perspective, the allele-specific real-time PCR assays performed well as demonstrated by the correct nucleotide identity being reported for each SNP. Allele-specific real-time PCR using SYBR green is appropriate for interrogation of SNPs because the method is reproducible, cost-effective and requires only standard laboratory equipment. In comparison with other real-time PCR detection methods such as TaqMan (Livak, 1999), the SYBR green does not require expensive, fluorescently-labelled oligonucleotide primers or probes. Costs can also be lowered by reducing the reaction volume, and by formulating criteria for determining the nucleotide identity through interrogation of a single allelic specific primer, which reduces the total reactions required (Huygens *et al.*, 2006). Furthermore, allele-specific real-time PCR is amenable to high throughput as automation of both the assay setup and reaction is possible (Huygens *et al.*, 2006). Overall, the allele-specific real-time PCR methodology was efficient at interrogating the high D SNP set with minimal expenditure of time and cost in comparison with DNA sequencing.

When conducting nosocomial or community population studies it is essential for major clones to be distinguished. From the perspective of genotyping the Australian MRSA collection into meaningful *in vitro* clonal groups, the seven high *D* SNP set was found to provide adequate resolving power. With regard to the prominent CA-MRSA clones, the ST 1, 30 and 93 strains were each genotyped individually using the high *D* SNPs. However, this was not without complications. The ST30 profile included the related but clinically distinct ST36, and despite ST93 being a singleton, the profile included an unrelated ST59. It is feasible to address these abnormalities via inclusion of additional SNPs, however this is an inefficient way of increasing genotyping power. Of the Australian HA-MRSA strains, the high *D* SNP set also performed well with regard to the MLST clonal boundaries. The globally disseminated ST239 is typically multiresistant and is frequently recovered from Australian hospitals. Even though it is closely related to the ST8 MRSA clone, it was easily differentiated using the high *D* SNPs. In addition to minor clones that were resolved, the major ST22 strain (UK EMRSA-15) also possessed a unique SNP profile.

During performance analysis of the high *D* SNPs, two small collections of unknown CA-MRSA from regional Queensland carriage studies were included (Peleg *et al.*, 2005; Vlack *et al.*, 2006). Results revealed that the MRSA carriage strains in the community were not dominated by a single stain, but rather, consisted of multiple strains. Overall, the *in vitro* results obtained from real world investigations together with those obtained by genotyping known isolates, indicate the suitability of the high *D* SNPs for mainstream genotyping applications.

It is unclear, however, how informative intra-clonal complex population structures as revealed by eBURST analysis are, particularly when considering clonal bacterial species. SLVs derived by mutation are very closely related to the clonal complex progenitor, and the highly stochastic nature of this process suggests that the population structure within clonal complexes, as revealed by eBURST analysis, is likely to be unreliable. For example, it is doubtful that the same isolates would be identified as clonal complex progenitors and SLVs if a different set of genes were chosen for MLST analysis. Evidence to support this contention has been reported by Robinson

and co-workers (Robinson *et al.*, 2005), who used two separate MLST schemes to analyse a set of *S. aureus* clonal complex 30 isolates. The two MLST methods identified different isolates as SLVs. Furthermore, since MLST and SNP genotyping are inherently restricted to assessing the core genome, it is obvious that variation within the accessory genome would strengthen the resolving power of these methods, particularly if the variation was related to significant phenotypic differences, rather than silent SNPs. These considerations form the basis of the rationale for our search of binary markers that add meaningful resolution to the high *D* SNP set efficiently.

While the concept of combining low and high resolution *S. aureus* genotyping methods is not novel, an integrated method targeting multiple, carefully selected loci is an attractive proposition. During this study we assessed a number of binary targets that had potential to increase SNP typing resolution by defining intra-SNP profile variation. The purpose of the binary targets was twofold, firstly, to differentiate strains from unrelated SNP profiles, and secondly, to subtype related strains within the SNP profiles. We first assessed virulence genes with a binary nature that was consistent with previous reports (Moore and Lindsay, 2001; Peacock *et al.*, 2002). Within our diverse collection of Australian MRSA it was found that the *pvl*, *tst*, *sdr*E and *cna* genes were informative binary targets that in combination increased the resolving power of the SNP genotyping methodology. When compared with the collection studied by Peacock and co-workers, it is clear that the proportion of virulence genes are influenced by the type of clonal complex tested, as shown by the increased proportion of the CA-MRSA associated *pvl* gene in our collection (Peacock *et al.*, 2002).

Although from the data it is possible to draw only limited conclusions regarding pathogenic potential, the binary virulence genes were most useful for identifying subtypes within the SNP profiles. Of the virulence genes tested, the *pvl* toxin gene is the most relevant clinically due to its strong association with CA-MRSA and necrotising pneumonia (Labandeira-Rey *et al.*, 2007). Screening results from this study reflect this association as we revealed *pvl* carriage in each of Australia's three major CA-MRSA clones. However, the *pvl* toxin was not carried uniformly by each strain within these SNP groups, which, for the purpose of subtyping the ST1 and

ST30 SNP profiles, was beneficial. The observed variation was not unexpected as bacteriophage transduction is responsible for *pvl* mobility. We also sought to determine whether variation in the actual *pvl* phage could improve genotyping resolution, however for the ST1 clone, no variation in phage type was detected (see Appendix Figs 1 and 2). Similarly, the *tst* gene is linked to serious disease and also exhibits a mobility mechanism; however its benefit to SNP typing was minimal as it was not widely distributed across the clonal groups. Also, the binary nature of the collagen adhesion gene, *cna*, has been reported previously (Peacock *et al.*, 2002). Although this gene is located chromosomally and is non-mobile, *cna* variation was able to subtype the ST5/73, ST8, ST78/88 SNP profiles successfully. While no obvious mobility mechanisms exist, the *cna* gene is located in proximity to an unusual "genomic islet", which may explain a previously lost mobility mechanism. Considering the small number of isolates in groups such as ST5/73, it remains to be seen whether *cna* is universally capable of differentiating these unrelated STs.

The *sdr*E gene is located within a paralogous gene family that also features *sdr*C and *sdr*D. Although *sdr*E does not display any obvious means of mobility, it is considered to be binary as a result of recombinational deletion or sequence evolution leading to PCR detection failure. The allelic state of *sdr*E is termed *bbp* and in our collection was exclusive to the SNP ST30 group. The lack of *sdr*E or *bbp* was identified in the ST1 and ST239 SNP groups, which permitted the differentiation of an additional subtype within each group. It has been reported that *sdr*E is more strongly associated with invasive isolates than *bbp*, however, ST30 appears to be one of the most successful and virulent strains circulating within Australia and is also *bbp* positive. Therefore it is prudent to acknowledge the multifactorial nature of *S. aureus* pathogenesis, which is influenced significantly by host factors. Overall, binary typing of the *pvl*, *cna* and *sdr*E*/bbp* genes in combination with the high *D* SNPs was an effective method because nine SNP genotypes were subtyped into a total of 21 combinatorial types.

The SCC*mec* element is a crucial variable component of the MRSA genome. Two approaches were chosen to identify SCC*mec* combinatorial genotyping targets, the first was to screen for associated plasmids, and the second was to characterise the

93

structural types via *ccr* and *mec* complex allotyping. Plasmids pT181, pUB110 and pI258 were particularly effective at subtyping SNP groups that included multi-resistant HA-MRSA. This was evident given the way the ST239 isolates were divided into four subtypes and the ability of pUB110 to differentiate the HA-MRSA ST36 from the CA-MRSA ST30 isolates. As HA-MRSA isolates circulate within sterile and therefore harsh nosocomial environments, acquisition of additional antimicrobial resistance mechanisms is essential for survival, so this result was not unexpected. Interestingly, the CA-MRSA ST1 clone also displayed SCC*mec* plasmid variation, as six isolates harboured pT181. Detection of these particular MGEs, which themselves are located within a larger MGE, proved to be suitable accessory targets for use in combination with the high *D* SNPs. In contrast to the majority of binary virulence genes, the SCC*mec* plasmids offer a correlation with phenotype. Therefore, addition of these antimicrobial resistance plasmids to a clonal genotyping method is beneficial as it provides a tentative indication of potential for resistance.

SCC*mec* characterisation via *ccr* and *mec* class allotyping has identified five major structural types that serve to define MRSA clones globally. Interestingly, each of the five types were recovered from our collection, in addition to novel subtypes. With regard to the CA-MRSA isolates, our results reflected the global trend of an association with the shortened SCC*mec* IV (2B) and recently discovered type V (5C). These variants are considered to be well suited to community acquisition as they are easily transferred and do not carry superfluous multi-resistance genes that are metabolically taxing. Of particular interest were five isolates within the CA-MRSA ST1 SNP group that were found to carry a SCC*mec* variant characterised by the unusual amplification of *ccr*1 in combination with *mec* class C. These isolates were also positive for the tetracycline resistant plasmid pT181, however whether this plasmid is carried by the SCC*mec* or not remains to be seen. As expected, the HA ST239 isolates carried the type III (3A) element, and the ST36 was further differentiated from ST30 via carriage of the type II (2A) element. Ultimately, three extra subtypes were resolved by characterisation of the core SCC*mec* structural components. The SCC*mec* element is an important component of MRSA genotyping strategies, and it is foreseeable that SCC*mec* genotyping methods with better resolving power would greatly enhance MRSA genotyping.

The final locus assessed as having potential for addition to the high D SNPs is the accessory genome regulator. The single instance of intra SNP profile variation was observed to differentiate ST59 from ST93. This could be useful information if the ST93 specific SNP was not implemented. However, this is recommended since CA *pvl* positive ST59 isolates have been reported internationally (Takano *et al.*, 2007).

In conclusion, this chapter has shown that interrogating the seven high D SNPs provides a suitable base method for a PHRANA MRSA genotyping strategy. It was also shown that combining seven binary targets provides a significant increase in overall genotype resolution whilst providing clinically relevant information. The approaches described were able to differentiate a diverse collection of Australian MRSA isolates containing nine SNP types into 21 subtypes. These subtypes are unambiguous, can be obtained easily and are consistent with previous studies. Interrogation of polymorphisms within the SCC*mec* and *agr* loci has the potential to increase resolution in certain circumstances, as does the addition of more MLST SNPs.

CHAPTER FOUR

SYSTEMATIC DERIVATION OF GENOTYPING TARGETS FOR THE STAPHYLOCOCCAL CASSETTE CHROMOSOME *MEC*

4.1 INTRODUCTION

S. aureus strains that express resistance to β-lactam antibiotics and have acquired the mobile SCC*mec* element are referred to as MRSA (Ito *et al.*, 1999; Katayama *et al.*, 2000). MRSA strains have progressively become established within hospitals globally (Panlilio *et al.*, 1992; Witte *et al.*, 1997), and are now being recovered from the community at an alarming rate (Vandenesch *et al.*, 2003). The full characterisation of MRSA strains requires the analysis of the core genome in concert with the SCC*mec* element; therefore SCC*mec* genotyping is conducted commonly in combination with a more general genotyping method.

The combination of SCC*mec* and a base genotyping method is a prime example of the PHRANA concept since both the background lineage and a variable region are combined to produce an informative genotype. Initial combination of PFGE and SCC*mec* genotyping demonstrated the importance of SCC*mec* characterisation as structural variation within and between the major HA-MRSA lineages was identified (Oliveira *et al.*, 2001). More recently, SCC*mec* typing in combination with MLST (Okuma *et al.*, 2002) and PFGE (Daum *et al.*, 2002) have provided evidence that indicates that CA and HA-MRSA clones had independent origins. Novel CA-MRSA SCC*mec* variants have been discovered and were found to be smaller in length and disseminated within diverse lineages by an undetermined mechanism. The high *D* SNP genotyping method described in Chapter One has the attributes of a clonal genotyping method that would be ideal for use with SCC*mec* genotyping. The apparent mobility of SCC*mec* within the species, and particularly within the general public, highlights the essential role of genotyping for understanding and managing the ever-increasing threat of antibiotic resistance.

While each SCC*mec* element features a *ccr* complex, a *mec* complex and non-essential junkyard regions, the characterised variants differ considerably in length and content. Overall, six SCC*mec* structural classes have been identified and are based on the combination of *ccr* and *mec* class (Ito *et al.*, 2001; Ito *et al.*, 2004; Ma *et al.*, 2002). In addition to the core SCC*mec* attributes, extensive variation is also evident due to insertion/deletion of plasmids, transposons and insertion sequences

(Chongtrakool *et al.*, 2006). These MGEs primarily insert within the junkyard regions of SCC*mec* and increase genomic diversity, however insertion sequences are also known to insert within the *mec* complex to create novel allotypes (Shore *et al.*, 2005). Recombination between SCC*mec* elements is also known to occur, which serves to increase SCC*mec* diversity (Lina *et al.*, 2006).

The emergence of CA-MRSA necessitates monitoring the spread of SCC*mec*, which has resulted in increased SCC*mec* genotyping output. This in turn, has led to the discovery of numerous novel elements with extensive diversity. Considering this, it is therefore appropriate that genotyping methods that can differentiate the myriad of known and potential types be developed and utilised. A number of broadly similar but distinct SCC*mec* typing methods have been described to date. These methods are based primarily upon PCR-based detection of type-specific targets including the *ccr* variants, *mec* gene complexes, and junkyard regions (Francois *et al.*, 2004; Okuma *et al.*, 2002; Oliveira and de Lencastre, 2002; Zhang *et al.*, 2005; Boye *et al.*, 2007) and have reported varying levels of success. These typing methods, and the associated terminology, have evolved in a somewhat *ad hoc* fashion as more SCC*mec* variants have been discovered. Recently, a proposal for a rationalised SCC*mec* typing nomenclature was published. This approach has merit since it incorporated the core structural features and the variable junkyard regions (Chongtrakool *et al.*, 2006). In an attempt to utilise much of the catalogued variation to date, an extensive SCC*mec* genotyping method was reported to have achieved high resolution, but required testing of 27 targets using six multiplex PCRs (Kondo *et al.*, 2007).

Despite this laborious approach, there is no current method that can harness the known SCC*mec* variation efficiently to achieve maximal resolution with minimal time and cost. Current methods are based around the principle of one target per genotype, which is inefficient and ultimately cannot accommodate the abundance of SCC*mec* variation. Although it is a difficult task to identify informative genotyping targets manually from complex loci such as the SCC*mec* element, it is possible through the use of computer software. Previously a computer program called Minimum SNPs was developed for the purpose of identifying SNPs from MLST databases with maximal resolving power (Robertson *et al.*, 2004). Resolving power

may be measured either by assessing the power to discriminate a user-defined sequence type from all other sequence types or on the basis of maximisation of the Simpson's index of diversity (D) (Hunter and Gaston, 1988).

More recently, the software was used to identify resolution-optimised sets of binary markers from a data set of binary marker variability. Instead of analysing actual MLST sequence data, binary marker variation was converted into a 'pseudo-DNA' sequence where the presence or absence of a gene was listed with a "T" or an "A", respectively. This pseudo sequence became the input for Minimum SNPs and was assessed to identify a set of binary markers that produce high genotyping resolution with respect to initial diversity. Using this approach, a set of such markers from *Campylobacter jejuni* comparative genome hybridisation data were identified on the basis of D maximisation, and these markers were shown to have considerable utility as genotyping targets (Price *et al.*, 2006).

SCC*mec* diversity can also be considered to be a database of binary marker variation and is therefore amenable to a similar analysis. However, in comparison to MLST comparative data, no such database of SCC*mec* variation exists. In fact, the variety of different genotyping methods used to characterise SCC*mec* serves to increase the difficulty of assessing binary marker diversity. Nevertheless, in order to support the central hypothesis of this chapter - that sets of markers derived by a systematic approach would provide resolving power efficiently - a data set of SCC*mec* binary marker data was constructed and Minimum SNPs analysis applied to it. Diversity within the SCC*mec*, however, is not restricted to allotypes of core genes, or junkyard associated binary markers. As the *mec*A gene is central to each SCC*mec* element, a secondary aim of this study was to determine if *mec*A SNPs could be utilised to add useful genotyping resolution. This was tested by concatenating actual DNA sequence to the binary marker pseudo sequence, then using Minimum SNPs to identify optimal combinations of genotyping targets.

4.2   MATERIALS AND METHODS

4.2.1   Collation of SCCmec Variation

An extensive literature and NCBI database search was undertaken to comprehensively catalogue the known SCC*mec* diversity in *S. aureus*. Previously unnamed variants were named according to the proposed SCC*mec* nomenclature system (Chongtrakool *et al.*, 2006). This procedure was complicated by the fact that described SCC*mec* variants differ in the detail to which they have been characterised; for some complete sequences are available, while others are classified only from PCR amplification. In a small number of instances, SCC*mec* variants reported in the literature lacked sufficient structural information to be included in this study.

4.2.2   Identification of Resolution-Optimised Sets of Binary Markers

Although Minimum SNPs can measure resolving power in several ways, the most generally applicable method is the calculation of D with respect to the input sequence alignment (Robertson *et al.*, 2004). This algorithm was used for this study. Resolution-optimised sets of binary markers were identified by first converting the binary marker data for the SCC*mec* variants into a string of "A's" and "T's," with "A" denoting binary marker absence and "T" denoting binary marker presence (Fig. 4.1). In this way, binary data for each SCC*mec* variant becomes a pseudo-DNA sequence that can be aligned with other pseudo-DNA sequences representing other SCC*mec* variants (Table 4.1). The alignment was then analysed using Minimum SNPs in order to identify sets of binary markers that give a high D value with respect to that alignment. This is achieved by identifying the single SNP with the highest resolving power, identifying it as SNP 1, then identifying the SNP that, in combination with SNP 1, gives the highest resolving power and labelling it SNP 2, and so on. A valuable feature of the software is that the user can force the program to include or exclude any SNP in/from the SNP set.
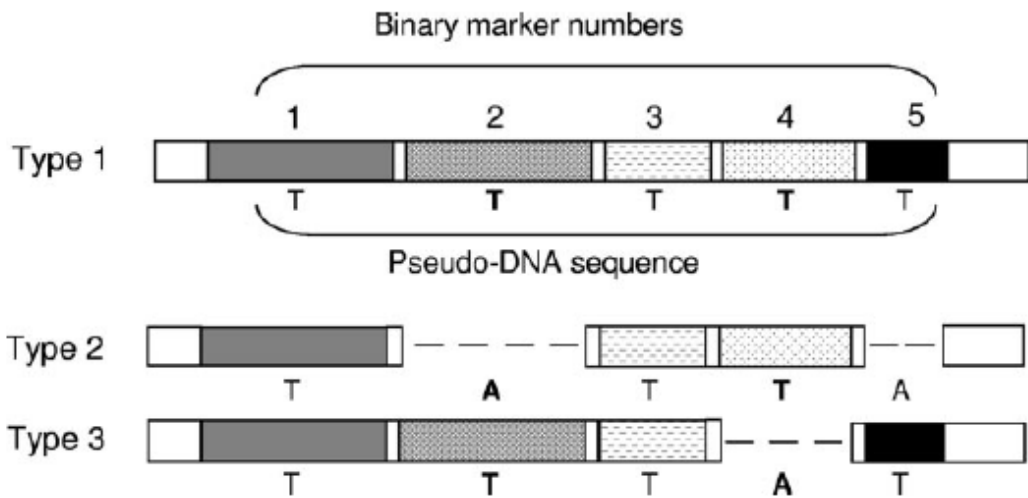
Fig 4.1. Illustration of the strategy used to identify sets of resolution-optimised binary markers. Each genotype of interest is characterised based on the presence (T) or absence (A) of the total set of binary markers (five are used in this example). Thus the binary marker configuration is converted into a pseudo-DNA sequence composed of A's and T's. The alignment of pseudo-DNA sequences is then analysed using Minimum SNPs for combinations of binary markers that maximize the Simpson's index of diversity (*D*) (Price *et al.*, 2006). In the example shown above, binary markers 2 and 4 provide a *D* of 1, that is, they completely resolve the three genotypes. In the present study, 46 SCC*mec* types were defined by the presence or absence of 34 binary markers. Even with this relatively small data set, manual identification of resolution-optimised marker sets is extremely difficult.

Table 4.1. Actual input data set reflecting SCC*mec* binary diversity

The rows below list the genetic markers; the columns list the SCC*mec* types. Cell values are A (absent) or T (present).

SCC*mec* type (columns, in order):
1B.1.1, (1B.2.1), 1B.1.2, (1B.1.3), (1B.1.4), 2A.1.1, 2A.1.2, (2A.13), 2A.3.1, 2A.3.2, 2A.3.3, 2A.3.4, 2A.3.5, 2A.2, (2F.1.1), 3A.1.1, (3A1.1.1), 3A.1.2, 3A.1.3, (3A.15), (3A.1.4), (3A.2.1), (3A.3.1), (3A.4.1), (3A.1.6), 2B.1, 2B.2.1, 2B.3.1, 2B.3.2, 2B.4, 2B.3.3, 2B.2.2, 2B.5, (2B.6), 4B, (2B.1.2), 2B.N.2, 5C.1, (5B.1), (5E.1), (5C2.1), (5/2G.1), (5/1C.1.1), (2C.1), Truncated

Genetic markers (rows):
ccr1, ccr2, ccr3, ccr4, ccr 5, ccrC-VT, mec class A, mec class A1, mec class A.3, mec class A.4, mec class B, mec class B1, mec class C2, mec class E, mec class F, mec class G, Jya, Jyb, Jyc, lvd, Jyg, pT181, pI258, pUB110, dcs, pls, kdp, Tn554MLS, tn554 CAD, Tn4001, 3A.2.1 unique, IS256-dcs, IS256-Tn4001, IS256-mecI

4.2.3 Identification of Resolution-Optimised Sets of Binary Markers for Abundant Structural Types

The alignment input file reflecting estimated SCC*mec* abundance included 49 additional copies of SCC*mec* variants: 1B.1.1 (I), 2A.1.1 (II), 3A1.1.1 (III), 3A.1.2 (IIIA), 3A.1.3 (IIIB), 2B.1 (IVa), 2B.2.1 (IVb), 2B.3.1 (IVc), and 5C.1 (V).

4.2.4 Calculation of Binary Target Set Performance and Generation of Reverse Identification Code

The genotyping performance of the binary targets with respect to the 46 SCC*mec* variants was assessed in Microsoft Excel. Briefly, binary target profiles corresponding to the chosen target set were compared visually to identify differences. Variants with shared profiles were considered identical in relation to the maximum resolution of the binary target set being assessed. The reverse identification code for the optimal target set was also devised in this manner.

4.2.5 mecA Nucleotide Sequence Determination

The *mec*A genes of 19 diverse Australian MRSA isolates from nine MLST types were amplified using primers *mec*AF1 and *mec*AR3 (Sigma-Proligo, Lismore, Australia) (Shukla *et al.*, 2004a). Amplicons were purified using Exo-SapIt (GE Healthcare Bio-Sciences Pty. Ltd., Sydney, Australia) for 15min at 37°C and 15min at 80°C and then sequenced for ~1,400bp from the 3' end using primers *mec*AF1 and *mec*AR2. The sequence traces were viewed and analysed in SeqMan II version 4.06 (DNAstar, Wisconsin, United States of America).

4.2.6   mecA Nucleotide Sequence Accession Numbers

Novel *mec*A sequences derived from this study were deposited in GenBank under accession numbers EF692630, EF692631, and EF692632.

4.2.7   Incorporation of mecA SNPs into SCCmec Binary Marker Input Data set

The literature and NCBI databases were searched for *S. aureus mec*A gene sequences where the corresponding SCC*mec* type was known. Twenty-five *mec*A sequences were identified, cropped to 1,440bp, and added to a sequence alignment with the 19 partial *mec*A sequences identified in this study. ClustalX (v. 1.64b, www.clustal.org) was used to align the partial *mec*A gene sequences to identify SNPs. In order to investigate the possibility of combinatorial genotyping methods based on SNPs and binary markers, actual SNP data was added to the pseudo-DNA sequences derived from the binary marker genotypes. This data became the input file for Minimum SNPs and allowed the identification of resolution-optimised SNP/binary marker combinations. Full details of the *mec*A sequences and isolates used in this analysis are found in Appendix Tables 2 and 3.

4.3 RESULTS

4.3.1 Collation of Available Data Concerning SCCmec Binary Diversity

The overall aim of this study was to apply a systematic approach for the identification of genetic targets that efficiently genotype SCC*mec*. In order to do this effectively, it was necessary to extract all available SCC*mec* data from the literature and public available databases. Diversity was available in the form of binary variability (gene presence/absence) or SNPs. As complete sequences were known for only a subset of the genes that make up all known SCC*mec* variants, binary diversity data were collated and analysed first.

A comprehensive search of the literature and DNA sequence databases resulted in identification of 34 SCC*mec*-associated binary targets (Table 4.2). The 34 binary targets in turn defined 46 SCC*mec* variants (Table 4.3). Each SCC*mec* element that differed at one or more of the 34 binary targets was considered a separate variant. This unprecedented compilation of SCC*mec* variants revealed that extensive rearrangement and mutation events had occurred during the evolution of the six major types, with diversity within the junkyard regions defining the majority of variants.

Table 4.2. The 34 binary targets used to define each SCC*mec* variant.

Core features		Junkyard features	
ccr	*mec* class	J1	J2/J3
*ccr*1	A	a	*dcs*
*ccr*2	A1[a]	b	*pls*
*ccr*3	A.3	c	*kdp*
*ccr*4	A.4	d	pT181
*ccr*5 (C)	B	g	pI258
*ccr*C-VT	B1[b]		pUB110
	C2		Tn*4001*
	E		Tn*554* (MLS)
	F[b]		ΨTn*554* (*cad*)
	G[c]		IS*256-dcs*
			IS*256*-Tn*4001*
			IS*256-mec*I
			3A.2.1 unique

[a] *mec* class A1 is characterised by a 166bp deletion in *mec*R1 (Ito *et al.*, 2001; Lim *et al.*, 2002; Oliveira *et al.*, 2000).

[b] Two separate *mec* classes have been termed B1. For this analysis, the *mec* class described by Lim *et al* remains B1 and the *mec* configuration described by Shukla *et al* has been renamed *mec* class F (Lim *et al.*, 2003; Shukla *et al.*, 2004a).

[c] The novel *mec* class featured in SCC*mec*ZH47 at the time of writing was unnamed, therefore the tentative name of *mec* class G was used (GenBank Accession No. AM292304).

Table 4.3. Forty-six SCC*mec* variants collated for this study[a]

SCC*mec* Class	SCC*mec* variant (*n* = 46)	Uniform Nomenclature [b]	*ccr*/*mec* class	J1	J2 and J3	PCR genotype [c]	Strain/isolate and/or Sequence (accession no.)	Reference(s)
I (1B)	I	1B.1.1	*ccr*1, *mec* class B	*pls*		I	NCTC10442 (AB033763), COL (CP000046)	(Ito *et al.*, 2001)
	I variant	(1B.2.1)	*ccr*1, *mec* class B	No *pls*		I	PhII	(Shore *et al.*, 2005)
	IA	1B.1.2	*ccr*1, *mec* class B	*pls*	pUB110	I	PER34	(Oliveira *et al.*, 2001)
	IA variant	(1B.1.3)	*ccr*1, *mec* class B	*pls*	IS256	I	PER184	(Oliveira *et al.*, 2001)
	IA variant	(1B.1.4)	*ccr*1, *mec* class B	*pls*	IS256 and pUB110	I	PER88	(Oliveira *et al.*, 2001)
II (2A)	IIa	2A.1.1	*ccr*2, *mec* class A	*kdp* operon	pUB110	II	N315 (D86934), MRSA252 (BX571856)	(Ito *et al.*, 1999)
	II variant	2A.1.2	*ccr*2, *mec* class A	*kdp* operon		II	Not described	(Cha *et al.*, 2005)
	II variant	(2A.1.3)	*ccr*2, *mec* class A	*kdp* operon	pUB110 and no *dcs*	II	Not described	(Cha *et al.*, 2005)
	IIA	2A.3.1	*ccr*2, *mec* class A.4 (IS *1182*)	J1 of IVb (2B.2.1)	Tn554 and pUB110	II NV	Not described	(Shore *et al.*, 2005)
	IIB	2A.3.2	*ccr*2, *mec* class A	J1 of IVb (2B.2.1)	pUB110	II	AJ810123	(Shore *et al.*, 2005)
	IIC	2A.3.3	*ccr*2, *mec* class A.3 (IS *1182*)	J1 of IVb (2B.2.1)	Tn554 and pUB110	II NV	Not described	(Shore *et al.*, 2005)
	IID	2A.3.4	*ccr*2, *mec* class A.4 (IS *1182*)	J1 of IVb (2B.2.1)	Tn554	II NV	Not described	(Shore *et al.*, 2005)
	IIE	2A.3.5	*ccr*2, *mec* class A.3 (IS *1182*)	J1 of IVb (2B.2.1)	Tn554	II NV	AR13.1/330.2 (AJ810120)	(Shore *et al.*, 2005)
	IIb	2A.2	*ccr*2, *mec* class A	Unique sequence for 2A.2	IS256, Tn554	II	JCSC3063 (AB127982)	(Hisata *et al.*, 2005)
	IIb variant	(2F.1.1)[d]	*ccr*2 , *mec* class F	*kdp* operon	*dcs* and no pUB110	NV	Not described	(Shukla *et al.*, 2004b)
III (3A)	III	(3A.1.1)	*ccr*3, *mec* class A		Tn554 MLS, pT181 and pI258	III	HUSA304	(Lim *et al.*, 2003; Oliveira *et al.*, 2001)
	III	(3A1.1.1)[e]	*ccr*3, *mec* class A1 (Δ*mec*R1)		Tn554 MLS, pT181 and pI258	III	85/2082 (AB037671), ANS46	(Ito *et al.*, 2001; Oliveira *et al.*, 2001)
	IIIA	3A.1.2	*ccr*3, *mec* class A		No pT181 or *ips*	III	HU25 (AF422651 to AF422696)	(Oliveira *et al.*, 2001)
	IIIB	3A.1.3	*ccr*3, *mec* class A		No pT181, pI258 or Tn554	III	HDG2	(Oliveira *et al.*, 2001)
	III variant	(3A.1.5)	*ccr*3, *mec* class A		No Tn554	III	R35	(Oliveira *et al.*, 2001)
	III variant	(3A.1.4)	*ccr*3, *mec* class A		pUB110	III	15814-9852	This study
	III variant	(3A.2.1)	*ccr*3, *mec* class A	Tn554 cad with *tnp*A		III	85/3907	(Ito *et al.*, 2003)
	III variant	(3A.3.1)	*ccr*3, *mec* class A	Tn554 (MLS)		III	85/961	(Ito *et al.*, 2003)
	III variant	(3A.4.1)	*ccr*3, *mec* class A	*pls*	*dcs*	III	DOM068	(Deurenberg *et al.*, 2005)
	III variant	(3A.1.6)	*ccr*3, *mec* class A		*dcs*	III	HSA10	(Aires de Sousa and de Lencastre, 2003)
IV (2B)	IVa	2B.1	*ccr*2, *mec* class B	Specific for 2B.1 (IVa)		IVa	MW2 (BA000033), CA05 (AB063172)	(Ma *et al.*, 2002)
	IVb	2B.2.1	*ccr*2, *mec* class B	Specific for 2B.2.1 (IVb)		IVb	8/6-3P (AB063173)	(Ma *et al.*, 2002)
	IVc	2B.3.1	*ccr*2, *mec* class B	Specific for 2B.3.1 (IVc)	Tn4001	IVc	MR108 (AB096217)	(Ito *et al.*, 2003)
	IVc variant	2B.3.2	*ccr*2, *mec* class B	Specific for 2B.3.1 (IVc)	No Tn4001	IVc	2314 (AY271717)	(Mongkolrattanothai *et al.*, 2004)
	IVd	2B.4	*ccr*2, *mec* class B	Specific for 2B.4 (IVd)		IVd	JCSC4469 (AB097677)	(Hisata *et al.*, 2005)
	IVE	2B.3.3	*ccr*2, *mec* class B	Identical to 2B.3.1 (IVc)	Unique left extremity sequence	IVc	AR43/3330.2 (AJ810121)	(Shore *et al.*, 2005)
	IVF	2B.2.2	*ccr*2, *mec* class B	Identical to 2B.2.1 (IVb)	Unique left extremity sequence	IVb	AR43	(Shore *et al.*, 2005)
	IVg	2B.5	*ccr*2, *mec* class B	Specific for 2B.5 (IVg)		IVb	M03-68 (DQ106887)	(Kwon *et al.*, 2005)
	IV variant	(2B.6)	*ccr*2, *mec* class B	Different to IVa,-b,-c or -d		IV NV	SD179-1	(Hisata *et al.*, 2005)
	IV variant	4B	*ccr*4, *mec* class B			NT	HDE288 (AF411935)	(Oliveira *et al.*, 2001)
	IV variant	(2B.1.2)	*ccr*2, *mec* class B		IS256	IVa	BARGII17	(Oliveira *et al.*, 2001; Roberts *et al.*, 1998)
	IVA variant	2B.N.2	*ccr*2, *mec* class B		pUB110	IVa	PER2	(Cha *et al.*, 2005; Oliveira and de Lencastre, 2002)
V (5C)	V	5C.1	*ccr*5, *mec* class C2	*hsd*		V	WIS (AB121219)	(Ito *et al.*, 2004)
	V variant	(5B.1)	*ccr*5, *mec* class B	No details described		NV	04-17489	(O'Brien *et al.*, 2005)
	V variant	(5B1.1)	*ccr*5, *mec* class B1	No details described		NV	WBG8404	(O'Brien *et al.*, 2005)
	V variant	(5E.1)	*ccr*5, *mec* class E	No details described		NV	WBG10198	(O'Brien *et al.*, 2005)
	VT	(5C2.1)	*ccr*C2, *mec* class C2	No details described		V	TSGH 17	(Boyle-Vavra *et al.*, 2005)
	V variant	(5/2G.1)	*ccr*5 and *ccr*2, *mec* class C2	*hsd*		NV	ZH 47 (AM292304)	Not published
	V variant	(5/1C.1.1)	*ccr*5 and *ccr*1, *mec* class C2	No details described	pT181	NV	B827549	This study
Other	NV	(2C.1)	*ccr*2, *mec* class C2	No details described		NV	04-16419	(O'Brien *et al.*, 2005)
	Truncated	-	-	-	IS431, pUB110 and *dcs*	NT	479968	(Donnio *et al.*, 2005)

ᵃ Table adapted from Chongtrakool and co-workers, with permission (Chongtrakool *et al.*, 2006).

ᵇ Uniform nomenclature values in parentheses denote tentative naming of unnamed variants as defined by Chongtrakool *et al.,* (Chongtrakool *et al.*, 2006).

ᶜ Hypothetical SCC*mec* genotype using standard PCR assays for *ccr*, *mec* class, and the 2B J1 regions a, b, c, and d. NV and NT indicate a new variant and a nontypeable variant, respectively.

ᵈ Shukla *et al.* described this isolate as a SCC*mec* 2A variant (Shukla *et al.*, 2004b).

ᵉ The *mec* class of isolate 85/2082 features a 166-bp deletion of the *mecR1* membrane spanning domain, which is characteristic for *mec* class A1 (Ito *et al.*, 2001; Lim *et al.*, 2002; Oliveira *et al.*, 2000).

4.3.2 Identification of a Resolution-Optimised Set of Binary Targets, Without Constraints

The database of binary marker variation was converted into a pseudo-sequence alignment and analysed using Minimum SNPs for sets of markers that provided a high D. In the first experiment, a set of binary markers were derived from the data in Table 4.1 simply on the basis of maximising D, with no attempt made to make the results consistent with previously described typing methods. All 46 SCC*mec* variants were completely resolved by interrogating 22 of the 34 binary targets. Table 4.4 shows the relative resolving powers of the best 6-, 9-, 12-, and 15-target sets that differentiate 21, 29, 35, and 39 of the 46 variants, respectively.

Table 4.4. Resolution-optimised sets of binary targets, selected without constraints

SCC*mec* class	SCC*mec* variants identified using a set of[a]:			
	6 targets (*ccr2*, Tn*554* MLS, *dcs*, pUB110, *mec* class A, and J1b)	9 targets (previous 6 plus *mec* class C2, IS256-*dcs*, and pT181)	12 targets (previous 9 plus *mec* class B, *ccr1*, and *mec* class A.3)	15 targets (previous 12 plus J1a, J1c, and *pls*)
1B (I)	1B.1.1	1B.1.1	1B.1.1	1B.1.1
	(1B.2.1)	(1B.2.1)	(1B.2.1)	(1B.2.1)
	4B*	4B*	4B	4B
	(1B.1.3)	1B.1.3)	(1B.1.3)	(1B.1.3)
	1B.1.2	1B.1.2	1B.1.2	1B.1.2
	Truncated*	Truncated*	Truncated	Truncated
	(1B.1.4)	(1B.1.4)	(1B.1.4)	(1B.1.4)
2A (II)	2A.3.1	2A.3.1	2A.3.1	2A.3.1
	2A.3.3	2A.3.3	2A.3.3	2A.3.3
	2A.3.4	2A.3.4	2A.3.4	2A.3.4
	2A.3.5	2A.3.5	2A.3.5	2A.3.5
	2A.1.1	2A.1.1	2A.1.1	2A.1.1
	2A.3.2	2A.3.2	2A.3.2	2A.3.2
	2A.2	2A.2	2A.2	2A.2
	2A.1.2	2A.1.2	2A.1.2	2A.1.2
	(2A.1.3)	(2A.1.3)	(2A.1.3)	(2A.1.3)
	(2F.1.1)	(2F.1.1)	(2F.1.1)	(2F.1.1)
3A (III)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)
	3A.1.2	3A.1.2	3A.1.2	3A.1.2
	3A.1.3	3A.1.3	3A.1.3	3A.1.3
	(3A.1.5)	(3A.1.5)	(3A.1.5)	(3A.1.5)
	(3A.1.6)	(3A.1.6)	(3A.1.6)	(3A.1.6)
	(3A.4.1)	(3A.4.1)	(3A.4.1)	(3A.4.1)
	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)
	(3A.1.4)	(3A.1.4)	(3A.1.4)	(3A.1.4)
2B (IV)	2B.4 (2B.6)	2B.4 (2B.6)	2B.4 (2B.6)	2B.4 (2B.6)
	2B.1	2B.1	2B.1	2B.1
	2B.3.2	2B.3.2	2B.3.2	2B.3.2
	(2B.1.2)	(2B.1.2)	(2B.1.2)	(2B.1.2)
	2B.3.1 2B.3.3	2B.3.1 2B.3.3	2B.3.1 2B.3.3	2B.3.1 2B.3.3
	(5/2G.1)*	(5/2G.1)*	(5/2G.1)	(5/2G.1)
	(2C.1)*	(2C.1)	(2C.1)	(2C.1)
	2B.2.1 2B.5	2B.2.1 2B.5	2B.2.1 2B.5	2B.2.1 2B.5
	2B.N.2	2B.N.2	2B.N.2	2B.N.2
	2B.2.2	2B.2.2	2B.2.2	2B.2.2
5C (V)	5C.1 (5C2.1)	5C.1 (5C2.1)	5C.1 (5C2.1)	5C.1 (5C2.1)
	(5B1.1) (5E.1)	(5B1.1) (5E.1)	(5B1.1) (5E.1)	(5B1.1) (5E.1)
	(5B.1)	(5B.1)	(5B.1)	(5B.1)
	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)
No. of variants resolved (D)	21/46 (0.9507)	29/46 (0.9768)	35/46 (0.9855)	39/46 (0.9923)

[a] Horizontal lines indicate discrimination by the respective sets of binary targets. The discriminatory power increases as additional targets are added, so the number of horizontal lines increases from left to right. SCC*mec* variants in parentheses were named in this study. Asterisks indicate variants grouped with unrelated variants.

4.3.3   Performance of the Resolution-Optimised Target Sets Compared to Current SCCmec Genotyping Targets

The resolution-optimised sets of binary targets were then compared with those used most commonly, with respect to their identity, resolving power, and whether the variants they define correspond with current SCC*mec* classification schemes. Current PCR-based SCC*mec* typing generally makes use of 10 targets (*ccr1*, *-2*, *-3*, and *-5*, *mec* classes A, B, and C, and three junkyard regions from type 2B) to assign SCC*mec* elements into five structural types. This approach also provides limited subtyping information. When tested against the data set containing 46 variants, the 10 targets could discriminate 18 genotypes. Comparatively, the first 10 targets derived on the basis of *D* maximisation discriminated 32 genotypes. This supported our contention that binary target identification on the basis of computerised *D* maximisation can provide a superior result to an *ad hoc* approach. In order to further test this idea, the resolving powers of randomly selected sets of markers were determined. This approach consistently produced much lower resolving powers than marker sets selected on the basis of *D* maximisation (data not shown), thus supporting the utility of our method for target selection.

4.3.4   A Resolution-Optimised Set of Binary Targets Nucleated by ccr1 and ccr5

 It would be desirable for any new SCC*mec* genotyping method to define genotypes consistent with the accepted terminology and current models concerning degree of relatedness among SCC*mec* variants. In other words, a genotyping method that fails to discriminate some pairs of distantly related SCC*mec* variants is of limited use. The genotypes defined by the *D* maximised marker set largely met this requirement. However, the six- and nine-target sets were unable to discriminate the unrelated 4B and truncated SCC*mec* from 1B (I) variants and also the 5/2G.1 and 2C.1 from 2B (IV) variants (Table 4.4). Accordingly, a second set of resolution-optimised targets was derived using the "Include" function in Minimum SNPs. This forces the program to include a chosen marker(s) in the set. In this instance, *ccr1* and *ccr5* were forced into the set, as addition of these markers can discriminate the variants not

112

discriminated by the original marker set. In this analysis, Minimum SNPs builds marker sets using the set of *ccr1* plus *ccr5* as a starting point, so in effect the derived marker sets are nucleated by *ccr1* and *ccr5*.

The derived marker sets were very similar to the unconstrained set and provided almost identical resolution power. The majority of incorrectly grouped variants were correctly assigned with this modification. The only instances of disparate variants being grouped together were the 2C.1 with 2B variants and 2B.N.2 with 2A variants when the six-target set was employed (Table 4.5).

Table 4.5. Resolution-optimised sets of binary targets nucleated by *ccr*1 and *ccr*5

SCC*mec* class	SCC*mec* variants identified using a set of[a]:			
	6 targets (*ccr1*, *ccr5*, Tn*554* MLS, *dcs*, pUB110, and *ccr2*)	9 targets (previous 6 plus J1b, *mec* class C2, and IS*256-dcs*)	12 targets (previous 9 plus *mec* class A, pT181, and *mec* class A.3)	15 targets (previous 12 plus *mec* class B, J1a, and J1c)
1B (I)	1B.1.1 (1B.2.1)	1B.1.1 (1B.2.1)	1B.1.1 (1B.2.1)	1B.1.1 (1B.2.1)
	(1B.1.3)	(1B.1.3)	(1B.1.3)	(1B.1.3)
	1B.1.2	1B.1.2	1B.1.2	1B.1.2
	(1B.1.4)	(1B.1.4)	(1B.1.4)	(1B.1.4)
2A (II)	2A.1.1	2A.1.1	2A.1.1	2A.1.1
	2A.3.1	2A.3.1	2A.3.1	2A.3.1
	2A.3.3	2A.3.3	2A.3.3	2A.3.3
	2A.3.4	2A.3.4	2A.3.4	2A.3.4
	2A.3.5	2A.3.5	2A.3.5	2A.3.5
	2A.1.2	2A.1.2	2A.1.2	2A.1.2
	(2F.1.1)	(2F.1.1)	(2F.1.1)	(2F.1.1)
	2A.3.2	2A.3.2	2A.3.2	2A.3.2
	2B.N.2*	2B.N.2	2B.N.2	2B.N.2
	2A.2	2A.2	2A.2	2A.2
	(2A.1.3)	(2A.1.3)	(2A.1.3)	(2A.1.3)
3A (III)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)	3A.1.1 (3A.2.1) (3A.3.1)
	3A.1.2	3A.1.2	3A.1.2	3A.1.2
	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)
	3A.1.3	3A.1.3	3A.1.3	3A.1.3
	(3A.1.5)	(3A.1.5)	(3A.1.5)	(3A.1.5)
	(3A.1.4)	(3A.1.4)	(3A.1.4)	(3A.1.4)
	(3A.1.6) (3A.4.1)	(3A.1.6) (3A.4.1)	(3A.1.6) (3A.4.1)	(3A.1.6) (3A.4.1)
2B (IV)	2B.1	2B.1	2B.1	2B.1
	2B.3.2	2B.3.2	2B.3.2	2B.3.2
	2B.4 (2B.6)	2B.4 (2B.6)	2B.4 (2B.6)	2B.4 (2B.6)
	2B.2.1 2B.5	2B.2.1 2B.5	2B.2.1 2B.5	2B.2.1 2B.5
	(2B.1.2)	(2B.1.2)	(2B.1.2)	(2B.1.2)
	2B.3.1 2B.3.3	2B.3.1 2B.3.3	2B.3.1 2B.3.3	2B.3.1 2B.3.3
	2B.2.2	2B.2.2	2B.2.2	2B.2.2
	(2C.1)*	(2C.1)	(2C.1)	(2C.1)
5C (V)	5C.1 (5C2.1)	5C.1 (5C2.1)	5C.1 (5C2.1)	5C.1 (5C2.1)
	(5B.1)	(5B.1)	(5B.1)	(5B.1)
	(5B1.1) (5E.1)	(5B1.1) (5E.1)	(5B1.1) (5E.1)	(5B1.1) (5E.1)
	(5/2G.1)	(5/2G.1)	(5/2G.1)	(5/2G.1)
	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)
Other	4B	4B	4B	4B
	Truncated	Truncated	Truncated	Truncated
No. of variants resolved (*D*)	18/46 (0.9391)	28/46 (0.9729)	34/46 (0.9836)	37/46 (0.9903)

[a] Horizontal lines indicate discrimination by the respective sets of binary targets. SCC*mec* variants in parentheses were named in this study. Asterisks indicate variants grouped with unrelated variants.

4.3.5   A Resolution-Optimised Target Set Derived from an Input File Reflecting the Abundant SCCmec Variants

Our approach to identifying resolution-optimised sets of genotyping targets presupposed that the database was an accurate surrogate of the real population relationships. If this were the case, then the resolving power of the targets applied to the database could constitute a useful measure of their resolving power on real isolate collections. This presupposition may, however, be simplistic, because all SCC*mec* variants are not equally abundant. This could lead to a difference between the *D* value calculated from the database and the *D* value obtained from actual collections of isolates. Thus for the *D* for actual collections of isolates to be maximised, markers that discriminate between abundant genotypes should be included preferentially in the marker set.

This issue was addressed by developing an input file for Minimum SNPs that contained multiple copies of abundant SCC*mec* variants. It was impractical to compile a comprehensive, accurate determination of relative abundances in nature; hence, a judgment was made as to which SCC*mec* variants were relatively abundant based on the published literature, and these variants each featured 50 times in the Minimum SNPs input file. Since a marker that discriminates between abundant variants should have a large impact on the *D* value, the derived marker set should discriminate abundant variants with high resolution. It was also predicted that the derived markers would still be effective for discriminating less-abundant variants from abundant variants and from each other.

The resolution-optimised marker sets from this exercise are shown in Table 4.6. The target sets identified are significantly different from the unconstrained and *ccr1*- and *ccr5*-nucleated target sets, but do not discriminate as many genotypes. As expected, the resolution-optimised marker set provided a performance superior to the unconstrained and *ccr1*- and *ccr5*-nucleated marker sets in resolving the nine SCC*mec* variants identified as abundant. There were, however, a small number of instances where the six-target set failed to discriminate rare SCC*mec* variants from other unrelated SCC*mec* variants. This was rectified partially by manually promoting

the 13th identified target, *ccr1*, into the six-target set. This discriminated 4B and the aberrant variant we have termed "truncated" from the abundant unrelated variant 1B.1.1. The resulting seven-target set is shown in the first column of targets in Table 4.6. This target set, however, did not discriminate 5/2G.1 and 2C.1 from the abundant unrelated variant 2B3.1. The 11-target set could, however, discriminate all unrelated variants.

To further test the efficacy of this approach, a separate data set containing only the nine abundant variants was assembled and analysed. Minimum SNPs calculated that six of the 34 targets were required to attain a *D* of 1. The six targets were very different from the marker set in Table 4.6 and provided poor resolving power with the entire data set, as demonstrated by numerous instances of unrelated variants failing to be discriminated (data not shown). Therefore, the approach of using the entire data set, with multiple copies of abundant genotypes, was considered to be superior.

Table 4.6. Resolution-optimised binary target sets derived from an input file reflecting SCC*mec* variant abundance

SCC*mec*	7 targets (*ccr1*, *dcs*, Tn*554* MLS, *ccr2*, *mec* class A, J1b, and pT181)	9 targets (previous 7 plus pUB110 and *mec* class B)	11 targets (previous 9 plus *mec* class C2 and J1a)	13 targets (previous 11 plus *pls* and IS*256-dcs*)	Abundant variant
		SCC*mec* variants identified using a set of[a]:			
1B (I)	1B.1.1	1B.1.1	1B.1.1	1B.1.1	Yes
	(1B.1.3)	(1B.1.3)	(1B.1.3)	(1B.1.3)	No
	(1B.2.1)	(1B.2.1)	(1B.2.1)	(1B.2.1)	No
	1B.1.2	1B.1.2	1B.1.2	1B.1.2	No
	(1B.1.4)	(1B.1.4)	(1B.1.4)	(1B.1.4)	No
2A (II)	2A.1.1	2A.1.1	2A.1.1	2A.1.1	Yes
	2A.1.2	2A.1.2	2A.1.2	2A.1.2	No
	2A.2	2A.2	2A.2	2A.2	No
	(2A.1.3)	(2A.1.3)	(2A.1.3)	(2A.1.3)	No
	2A.3.2	2A.3.2	2A.3.2	2A.3.2	No
	(2F.1.1)	(2F.1.1)	(2F.1.1)	(2F.1.1)	No
	2A.3.1	2A.3.1	2A.3.1	2A.3.1	No
	2A.3.3	2A.3.3	2A.3.3	2A.3.3	No
	2A.3.4	2A.3.4	2A.3.4	2A.3.4	No
	2A.3.5	2A.3.5	2A.3.5	2A.3.5	No
3A (III)	3A.1.1	3A.1.1	3A.1.1	3A.1.1	Yes
	(3A.2.1)	(3A.2.1)	(3A.2.1)	(3A.2.1)	No
	(3A.3.1)	(3A.3.1)	(3A.3.1)	(3A.3.1)	No
	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)	(3A1.1.1)	No
	3A.1.2	3A.1.2	3A.1.2	3A.1.2	Yes
	(3A.1.4)	(3A.1.4)	(3A.1.4)	(3A.1.4)	No
	3A.1.3	3A.1.3	3A.1.3	3A.1.3	Yes
	(3A.1.5)	(3A.1.5)	(3A.1.5)	(3A.1.5)	No
	(3A.1.6)	(3A.1.6)	(3A.1.6)	(3A.1.6)	No
	(3A.4.1)	(3A.4.1)	(3A.4.1)	(3A.4.1)	No
2B (IV)	2B.1	2B.1	2B.1	2B.1	Yes
	(2B.1.2)	(2B.1.2)	(2B.1.2)	(2B.1.2)	No
	2B.3.2	2B.3.2	2B.3.2	2B.3.2	No
	2B.4	2B.4	2B.4	2B.4	No
	(2B.6)	(2B.6)	(2B.6)	(2B.6)	No
	2B.N.2	2B.N.2	2B.N.2	2B.N.2	No
	2B.2.1	2B.2.1	2B.2.1	2B.2.1	Yes
	2B.5	2B.5	2B.5	2B.5	No
	2B.3.1	2B.3.1	2B.3.1	2B.3.1	Yes
	2B.3.3	2B.3.3	2B.3.3	2B.3.3	No
	(5/2G.1)*	(5/2G.1)*	(5/2G.1)	(5/2G.1)	No
	(2C.1)*	(2C.1)*	(2C.1)	(2C.1)	No
	2B.2.2	2B.2.2	2B.2.2	2B.2.2	No
5C (V)	5C.1	5C.1	5C.1	5C.1	Yes
	(5C2.1)	(5C2.1)	(5C2.1)	(5C2.1)	No
	(5B1.1)	(5B1.1)	(5B1.1)	(5B1.1)	No
	(5E.1)	(5E.1)	(5E.1)	(5E.1)	No
	(5B.1)	(5B.1)	(5B.1)	(5B.1)	No
	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)	(5/1C.1.1)	No
Other	4B*	4B	4B	4B	No
	Truncated*	Truncated	Truncated	Truncated	No
No. of variants resolved	19/46 (0.8976)	28/46 (0.9012)	31/46 (0.9034)	36/46 (0.9047)[b]	

[a] Horizontal lines indicate discrimination by the respective sets of binary targets. SCC*mec* variants in parentheses were named in this study. Asterisks indicate variants grouped with

unrelated SCC*mec* types. [b] Due to the input data set containing repeats of identical sequences, a *D* value close to 1 cannot be obtained.

4.3.6   Correlation of Binary Target Profile with SCCmec Type

Determining the precise identity of SCC*mec* elements using Minimum SNPs-derived binary targets can be achieved in two ways. The first requires a working knowledge of Minimum SNPs, and is based on inputting the binary target profile of interest and working backward to obtain, from the original alignment, the variants that share the entered profile. This method is proportionately more useful in relation to an increasing number of possible result outcomes from the binary target set in question. The second process for determining a genotype is to construct a reference table that correlates the binary marker profile to genotype name. This process was feasible in this study, as the number of result outcomes possible was capped at a maximum of 46. As an example, this process allowed rapid correlation of binary target results from the seven abundance targets (Table 4.6) with a genotype (Table 4.7).

Table 4.7. Correlation between binary target profiles and SCC*mec* genotypes.

SCC*mec* binary profile[a]	Undifferentiated SCC*mec* types
AAAAAAA	5C.1*, (5C2.1), (5B1.1), (5E.1), (5B.1)
AAAATAA	3A.1.3*
AAAATAT	(3A.1.5)
AAATAAA	2B.3.1*, 2B.3.3, (5/2G.1), (2C.1)
AAATATA	2B.2.2
AATAAAT	(3A1.1.1)
AATATAA	3A.1.2*, (3A.1.4)
AATATAT	3A.1.1, (3A.2.1), (3A.3.1)
AATTTAA	2A.2, (2A.1.3)
ATAAAAA	4B, Truncated
ATATAAA	2B.1*, (2B.1.2), 2B.3.2, 2B.4, (2B.6), 2B.N.2
ATATATA	2B.2.1*, 2B.5
ATATTTA	2A.3.2
ATTATAT	(3A.1.6), (3A.4.1)
ATTTAAA	(2F.1.1)
ATTTATA	2A.3.1*, 2A.3.3, 2A.3.4, 2A.3.5
ATTTTAA	2A.1.1*, 2A.1.2
TAAAAAT	(5/1C.1.1)
TTAAAAA	1B.1.1*, (1B.1.3), (1B.2.1), 1B.1.2, (1B.1.4)

[a] First six abundance targets with *ccr*1, in order: *ccr*1, *dcs,* Tn*554* MLS, *ccr*2, *mec* class A, JYb, pT181

* Defines the nine abundant SCC*mec* types.

4.3.7 Binary Target Unique Sequence for Potential Real-Time PCR Assays

For targets identified here to be converted into real-time PCR-based assays, the identification of target specific sequence is required. Table 4.8 provides the location of sequence fragments specific for the seven targets of the abundance set (Table 4.7). These regions can be amplified to produce binary marker data profiles that efficiently genotype the listed SCC*mec* types to the resolution specified in Table 4.6.

Table 4.8. Unique sequence loci for the seven member abundance target set

First six abundance targets with *ccr*1	NCBI reference	Characteristic nucleotide sequence
Cassette chromosome recombinase A1 (*ccr*A1)	AB033763	23692..25041
dcs	AB063172	23341..24300
Tn*554* MLS	AJ810120	10361..11040
Cassette chromosome recombinase A2 (*ccr*A2)	BA000017	65008..66357
mec class A	BA000017	46999..49226
Jyb (Junkyard b or J1 region of 2B.2.1)	AB063173	2195..2310
pT181	AB037671	29641..30000

4.3.8   mecA SNPs Can Increase Resolving Power

It has been reported previously that the SCC*mec mec*A gene contains several SNPs (Ryffel *et al.*, 1992; Shukla *et al.*, 2004a; Wu *et al.*, 1992). These are potential genotyping markers that could be used either as replacements for the binary targets selected, or to define more SCC*mec* subtypes. The literature and sequence databases were searched for *mec*A sequences from characterised SCC*mec* elements. In total, 25 sequences were retrieved from the NCBI database, and in addition, partial *mec*A sequence analysis was undertaken on 19 selected isolates from our collection of characterised SCC*mec* elements. From this set of 44 partial *mec*A sequences, eight SNPs in total were identified, three of which were novel (Table 4.9).

Table 4.9. SCC*mec* variants and associated *mec*A SNP profiles [a]

SCC*mec* Class	SCC*mec* Variant	*mec*A SNP position							
		75	312*	415*	438*	448*	612*	675*	737
1B (I)	1B.1.1	A	C	G	T	G	T	T	A/G
2A (II)	2A.1.1	A/C	C	G	T	G	T	T	G
	2A.3.5	A	C	G	T	G	T	T	G
2B (IV)	2B.1	A	C	G	T	G	T	T	G
	2B.2.1	A	C	G	T	G	T	T	G
	2B.3.1	A	C	G	T	G	T	T	A/G
	2B.3.2	A	C	G	T	G	T	T	A
	2B.3.3	A	C	G	T	G	T	T	G
	2B.5	A	C	G	T	G	T	T	A
3A (III)	3A.1.1	A	C	G	T/A	G/A	G	T	A
	3A.1.2	A	A	G	T	G	G	T	A
	3A.1.3	A	C	G	T	G	T	T	G
	3A.1.4	A	C	G/A	T	G	T	T	G
	3A.2.1	A	C	G	T	G	T	T	A
5C (V)	5/1C.1.1	A	C	G	T	G	T	A	G
	5C.1	A	C	G	T	G	T	A	G
	5C2.1	A	C	G	T	G	T	A	G
	5/2G.1	A	C	G	T	G	T	A	G

[a] The most abundant SNP profile, ACGTGTTG, was found 22 times. *mec*A SNP references: positions 75 (Ito *et al.*, 1999), 312 (this study), 415 (this study), 438 (Taneike *et al.*, 2006), 448 (this study), 612 (Shukla *et al.*, 2004a), 675 (Ito *et al.*, 2004), and 737 (Shukla *et al.*, 2004a).

[b] Asterisks indicate non-synonymous codon changes.

To determine whether *mec*A SNPs add resolution to the binary markers, they were included in a Minimum SNPs analysis as additional binary data points. Because no precise *mec*A sequence data could be assigned to many of the SCC*mec* variants, a new SCC*mec* data set was created containing only variants with corresponding *mec*A sequences. In this case, the data used for input into Minimum SNPs was in effect an alignment containing a region of pseudo-DNA sequence derived from binary marker variation (A and T) and a region of actual DNA sequence derived from *mec*A sequences. This data set consisted of 18 of the original 46 SCC*mec* variants (Table 4.1 and 4.3). As different *mec*A SNP profiles were found within identical SCC*mec* variants, the 34 binary markers in combination with the eight SNPs defined 24

variants. These expanded and mixed SNP/binary target profiles were analysed using Minimum SNPs. This analysis resulted in a *D* of 1.0 using nine binary targets and five *mec*A SNPs. Overall, of the five SNPs, only SNP 737 could be considered a possible replacement for the binary targets, as it was selected at the third position in the Minimum SNPs output. The remaining SNPs were selected at positions 9 and 12 to 14 and demonstrated minimal *D* value contribution. The entire marker set includes; *ccr2*, Tn*554* MLS, *mec*A*737*, *mec* class A, *dcs*, pT181, *ccrC*, Jyb, *mec*A*438*, *ccr*C-VT, Tn*4001*, *mec*A*75*, *mec*A*415*, and *mec*A*448*.

4.4 DISCUSSION

It is now generally accepted that a powerful strategy for bacterial genotyping is to interrogate the genome backbone plus one or more hypervariable regions. This generalised approach has been referred to as "phylogenetic hierarchical assays using nucleic acids" or PHRANA (Keim *et al.*, 2004). The practice of identifying MRSA clones using both MLST and SCC*mec* type is an example of this. Widespread adoption of this MRSA genotyping strategy has given impetus to the accumulation of considerable information about SCC*mec* diversity and the development of several specific SCC*mec* subtyping schemes.

SCC*mec* has been classified as six general types based on *mec* classes and *ccr* gene identity. Several SCC*mec* genotyping methods classify the element to this level only (Francois *et al.*, 2004; Yang *et al.*, 2006). The results of a comprehensive search of the literature and sequence databases emphasised the high diversity of SCC*mec* and revealed that much of the diversity is invisible to previously published SCC*mec* typing methods. In addition to initial characterisation of types 1B, 2A, and 3A (Ito *et al.*, 2001), community-acquired type 2B (IV) (Ma *et al.*, 2002), and, most recently, type 5C (V) (Ito *et al.*, 2004), further variable and unusual variants have been identified.

Of particular interest are variants carrying multiple copies of *ccr* and instances of new combinations of *mec* and *ccr* classes (Boyle-Vavra *et al.*, 2005; O'Brien *et al.*, 2005). A recently identified complication for SCC*mec* genotyping is the discovery of additional SCC elements integrating in *orfX* adjacent to SCC*mec* that are identified through carriage of multiple *ccr* complexes. The prime example of this is the composite element constructed of SCC*mec* 3A and SCC*mer*. For the analysis detailed here, MRSA isolates suspected of carrying composite elements were included, as one of the goals was to enhance genotyping resolution power.

The approach adopted here of identifying sets of binary markers for genotyping SCC*mec* differs greatly from previously published methods (Kondo *et al.*, 2007). Genotyping approaches reported to date approximately yield only one *mec* class or *ccr* gene per marker interrogated. This is inefficient since recombination in SCC*mec*

means that resolving power could be increased logarithmically rather than arithmetically if more targets were interrogated. Accordingly, instead of identifying binary markers diagnostic for particular classes/types/subtypes, we used a computerised approach that identifies sets of markers that maximise D. This algorithm attempts to identify a marker that splits the known variants into two equal halves and then attempts to find a single marker that splits each of the groups defined by marker 1 into two equal halves, and so forth. In effect, it is a search for markers that are maximally unlinked. This approach has proven to be valuable. It has provided sets of markers with greater resolving power than equivalently sized sets of markers identified by traditional approaches. Also, the analyses clearly defined a range of options with respect to the numbers of markers interrogated and the resolution obtained.

The markers identified have potential to inform the design of specific SCC*mec* genotyping methods with respect to the resolution required and/or the multiplexing capacity of the technology to be used**.** Of particular interest is the strong consistency observed between the SCC*mec* classes and the groups defined by small numbers of resolution-optimised marker sets, even when there were no constraints on marker selection. The fact that maximally unlinked markers define these groups supports the notion that the SCC*mec* classes indeed represent distinct phylogenetic clusters of this element. The strategy to increase the consistency between genotypes and the relationships between the SCC*mec* variants by nucleating the marker set with *ccr1* and *ccr5* proved successful, as was the attempt to identify markers especially efficient at discriminating the abundant SCC*mec* variants. Overall, our approach for marker selection was effective and flexible. The marker sets identified provide a wide choice of well-understood options for the design of SCC*mec* genotyping procedures. This general approach could be applied to any genome region that displays a high degree of binary variability, e.g., the loci encoding the enzymes that synthesise complex, antigenically active, cell wall-associated polysaccharides.

The value of the systematically derived binary markers is proportional to the extent of variation included in the input data set. The optimal situation is composition of accurately detailed data sets that represent a wide gamut of variation for the loci in

question. For SCC*mec* this was somewhat difficult due to the number and resolution of genotyping methods used to identify the diversity reported to date. Of particular concern is that variants that have not been thoroughly characterised may contain configurations of key targets that were not previously known to exist. Unfortunately, certain novel SCC*mec* types were omitted from this analysis due to a lack of characterised features. Furthermore, certain binary targets were not assessed within the confines of this study due to insufficient data, including junction sequences, *ccr* sequence mutations and other junkyard features. It appears plausible that additional variants and binary targets could be included in future Minimum SNPs analyses for the purpose of improving the resolution of target sets.

Comprehensive SCC*mec* characterisation and nomenclature assignment require the unambiguous determination of *mec* and *ccr* complexes in addition to the junkyard regions. Rapid SCC*mec* genotyping with the aim of augmenting a clonal genotyping method, however, does not require such rigour. This observation helps to counteract concern regarding use of abbreviated methods such as the multiplex method used by Oliveira and co-workers and the method described here. Conveniently, optimal combinations of SCC*mec* targets identified here accommodate the beneficial features from both the simplex and multiplex methods. In fact, four of the first seven targets from the smallest abundance set are standard SCC*mec* genotyping targets. In contrast to the multiplex method, the junkyard targets were selected based on maximal combinatorial power rather than by simply indicating a single type. Overall, although the smaller Minimum SNPs target sets are not highly suited for complete SCC*mec* characterisation, they are ideal for the generation of rapid and efficient MRSA genotypes.

The effect of including *mec*A SNPs in the marker set was explored, and it was found that such SNPs defined additional SCC*mec* variants. These SNPs are not, however, effective substitutes for binary markers. Of interest was the observation that identical SNP allelic combinations were found in different SCC*mec* variants while in contrast, different *mec*A alleles were found in identical SCC*mec* variants. For example, the first four SCC*mec* structural types (1B, 2A, 3A, and 2B) each carry the most abundant

*mec*A sequence, while seven of the nine *mec*A sequences are found in type 3A. This suggests that some SCC*mec* types are much older than the others or that SCC*mec* recombination may have occurred. Other observations of interest included; variant 2B.1 carrying the most numerous *mec*A sequence exclusively, which is different from the *mec*A sequence in the closely related variant 2B.3.1 (IVc), and complete linkage found between *ccr5* and a particular *mec*A sequence (Table 4.9). It was also observed that two SNP profiles were primarily associated with *mec* class B, while six SNP profiles were associated with *mec* classes A. Thus, *mec*A SNPs may constitute a possible means for increasing resolving power by increasing the number of genotypes. However, the *mec*A SNPs have only a limited potential use as replacements for the binary markers, especially considering that it is more technically straightforward to assay for the presence of a binary marker than to interrogate the allelic identity of a SNP.

We used Minimum SNPs initially to derive sets of seven resolution-optimised SNPs from the *S. aureus* MLST database. Genotypes defined by the high D SNP set were highly consistent with *S. aureus* clonal complex structure determined by eBURST analysis of the MLST database. By interrogating these SNPs as well as, for example, the seven targets in Table 4.7 (*ccr1*, *dcs*, Tn*554* MLS, *ccr2*, *mec* class A, J1b, and pT181) we have identified a potential method for large scale MRSA clone identification. This method would provide sufficient resolution for many purposes and would be easily adapted to real-time PCR, medium-density array, or "lab on a chip" technologies. Additional resolution for exploring small-scale epidemiological questions (e.g. infection control) could be provided by interrogating one or more variable-number tandem repeat loci. This could also be done in a real-time PCR device by T_m measurement.

In conclusion, the results here demonstrate the effectiveness and flexibility of a systematic computer-assisted approach to extricate genotyping targets from loci that display high levels of complex binary marker variation.

CHAPTER FIVE

HIGH-RESOLUTION MELTING ANALYSIS OF THE *SPA* VARIABLE NUMBER OF TANDEM REPEAT REGION

5.1 INTRODUCTION

The development rate of molecular genotyping methods has increased significantly over the last decade. Although PFGE is considered the gold standard for *S. aureus* genotyping, due to issues relating to utility, experimental turn-around-time, and inter-laboratory comparability, it is likely to be replaced by novel methods derived from comparative genomics. MLST and MLST-based SNP typing are such methods that provide portable genetic fingerprints and produce practical clonal population structures. Although powerful for long-term investigations, these methods suffer from limited short-term resolution as they only consider the core genomic backbone of target species (Maiden *et al.*, 1998). Therefore, optimal genotyping will rely on a combined approach where targets from both the clonal and variable regions are surveyed. Genotyping the SCC*mec* element in combination with MLST has become more common for MRSA; however, there remains a need for higher resolution genotyping of both MSSA and MRSA so as to produce a competitive alternative to PFGE. This resolution is achievable by harnessing the informative power of highly polymorphic loci.

An ideal third tier component for a PHRANA based genotyping strategy would target maximally unlinked genomic loci that exhibit higher evolution rates than the prior methods. The standout candidates within the *S. aureus* genome are genes that encounter length and sequence variation from different numbers of repeat units. These regions are most commonly known as variable number of tandem repeats (VNTR) loci (Nakamura *et al.*, 1987). VNTR loci contain complex repeat configurations that are thought to mutate rapidly via slipped-strand mispairing and/or recombination (Brigido Mde *et al.*, 1991; Guss *et al.*, 1984; Uhlen *et al.*, 1984). Hyper-mutation allows VNTR loci to be extremely informative markers for discriminating among closely related bacterial strains, which is demonstrated by their successful implementation in genotyping strategies for *Bacillus anthracis* (Keim *et al.*, 2000) and *Escherichia coli* (Noller *et al.*, 2003).

In *S. aureus*, VNTRs are associated mainly with membrane-bound proteins classified as "microbial surface components recognising adhesive matrix molecules" (MSCRAMMs). The DNA tandem repeats are located between the conserved cell wall-spanning and protein-binding domains, and once translated act as a stalk to enable exposure of the active protein-binding domain to the extracellular environment (van Belkum *et al.*, 1998a). The extracellular exposure of the translated repeat region domain is subject to immune surveillance that is also expected to increase their rate of evolution. The MSCRAMM group include proteins such as fibrinogen-binding clumping factor B (*clfB*) (Ni Eidhin *et al.*, 1998), serine-aspartate repeat proteins (*sdr*C, D, E) (Josefsson *et al.*, 1998) and others (van Belkum *et al.*, 1998a). The most well studied *S. aureus* VNTR locus, however, is located within the Staphylococcal protein A (*spa*) gene (Lofdahl *et al.*, 1983). It has been suggested that protein A conceals *S. aureus* from polymorphic neutrophils by coating the cell surface with bound IgG molecules, thus preventing immune recognition (Uhlen *et al.*, 1984).

Genotyping methods that utilise VNTR variation are designed around PCR amplification from the conserved flanking regions across the repeat region, followed by agarose gel electrophoresis size determination or DNA sequencing (Shopsin *et al.*, 1999). The most established *S. aureus* VNTR genotyping method involves sequence analysis of the polymorphic region X from the *spa* gene (Harmsen *et al.*, 2003; Koreen *et al.*, 2004; Shopsin *et al.*, 1999). The *spa* 3` coding region contains 1-23 repeat units of 24bp that undergo a complex and rapid evolutionary process of deletion, duplication and point mutation (Brigido Mde *et al.*, 1991; Guss *et al.*, 1984; Uhlen *et al.*, 1984). The success of this typing method has been attributed to the extensive variation observed thus far that is recorded at http://spaserver.ridom.de/. To date, 2331 *spa* types featuring 155 different repeat units have been reported from ~30,000 isolates.

While *spa* sequence typing has been used successfully for both short-term and long-term epidemiological investigations, it has been shown to perform well against PFGE in regard to long-term discriminatory power and genotype concordance. There are instances, however, where MLST genotype classifications are not adhered to (Hallin

et al., 2007). A high rate of variation is beneficial for short-term investigations; however it is not desirable for long-term investigations as it can obscure the location of a particular strain in the overall population structure. In order to fully harness discriminatory power and to avoid homoplasy (the independent formation of identical genotypes), a coupling with a core genotyping method will be required. This concept is defined by the PHRANA model that enables high resolving power via the analysis of separate genomic loci that evolve at differing rates, while maintaining their clonal identity (Keim *et al.*, 2004). MLST and *spa* typing form such a partnership that is commonly employed; however, both methods are hindered by the time consuming and costly process of DNA sequencing.

With the aim of creating rapid, cheap and robust genotyping methods, we have previously used allele-specific real-time PCR to genotype *S. aureus* into MLST clonal groups. Therefore, addition of a real-time PCR compatible assay that utilises the diversity of a hypervariable region such as the *spa* locus is highly desirable. Although sequencing technology is now extremely effective and cost is declining, non–sequence-based genotyping methods maintain clear advantages. In particular, a real-time PCR platform supports genotyping methods that are single step, closed tube, can potentially be carried out simultaneously with diagnosis, and can interrogate different classes of genetic polymorphisms. These features provide real advantages for the clinical microbiology laboratory in regard to implementation of a PHRANA genotyping strategy.

A recent development in real-time PCR technology is high-resolution DNA melting (HRM) analysis (Gundry *et al.*, 2003; Wittwer *et al.*, 2003). Although DNA melting curves are used predominantly to determine the melting temperature (T_m) of amplified double-stranded DNA, the precise shape of a melting curve is a function of the DNA sequence, and this characteristic forms the basis of HRM analysis (Ririe *et al.*, 1997). Accurate melting curves are produced using very small temperature increments, whilst normalisation and comparison of melting curves allows for the sensitive determination of whether different amplicons have the same or a different sequence (Wittwer *et al.*, 2003). The potential resolving power of this approach is much greater than conventional T_m based melting curve analysis because HRM

melting curves from different amplicons can be differentiated on the basis of shape, even when they define the same T_m values.

HRM has been applied previously to human mutation screening (Krypuy *et al.*, 2006; Vaughn and Elenitoba-Johnson, 2004; Dufresne *et al.*, 2006) and has been implemented to differentiate the hypervariable *CRISPR* (clustered regularly interspaced short-palindromic-repeat) locus of *C. jejuni* (Price *et al.*, 2007). We tested the hypothesis that HRM analysis on the Corbett Rotor-Gene 6000 real-time PCR instrument can be used to differentiate *spa* alleles. We also hypothesised that HRM curves can be compared between different runs on the Rotor-Gene 6000, a characteristic that would indicate that HRM analysis can be used as a library-based genotyping method instead of the approach being strictly comparative.

5.2 MATERIALS AND METHODS

5.2.1 Bacterial Isolates

Forty-two well characterised Australian and two French MRSA isolates from 13 MLST clones were used in this study (Table 5.1). The core genotypes of each isolate had been determined previously using either MLST or SNP-based genotyping (Coombs *et al.*, 2004). To maximise the probability of identifying a spectrum of *spa* types, isolates were selected from diverse HA and CA MLST/SNP types. Additionally, multiple isolates from identical MLST/SNP types were selected to identify closely related *spa* types.

5.2.2 PCR Amplification

PCR primers for amplification and sequencing of the *spa* locus were designed previously and are as follows: 1095F (5'-ACGATCCTTCGGTGAGC-3') and 1517R (5'-GCTTTTGCAATGTCATTTACTG-3') (Shopsin *et al.*, 1999). Each reaction contained: PCR buffer (10x), 1.5mM $MgCl_2$, 0.2mM dNTPs, 0.5mM each primer, 0.25U platinum *Taq* polymerase (Roche Diagnostics, New South Wales, Australia), about 25ng DNA template and made up to 25μL with ddH_2O. Amplification was achieved using the following thermocycling protocol: 10min at 95°C; 30 cycles of 30s at 95°C, 30s at 60°C, and 45s at 72°C; and a final extension at 72°C for 10min. Resulting amplicons were subjected to 1.2% agarose gel electrophoresis using TBE buffer and ethidium bromide for 25min, at 120V.

5.2.3 DNA Sequencing

The *spa* repeat region of each isolate was sequenced using the PD service provided by the AGRF and each sequence chromatogram was analysed using Seqman by DNAstar.

133

5.2.4   spa Genotype Determination

Unlike the MLST database, there is no free method for automatically searching the http://spaserver.ridom.de/ database to assign *spa* types to an unknown sequence. Therefore, sequence data for each known repeat unit and *spa* type were imported into a Microsoft Excel spreadsheet and the "vlookup" function used to determine the identity of unknown repeat unit sequences and *spa* genotypes. Once each repeat unit had been assigned a corresponding name, the overall profile consisting of each repeat unit for the isolate in question was then assigned a *spa* genotype.

5.2.5   Real-Time PCR and High Resolution Melting Analysis

The conventional PCR *spa* assay was modified and optimised for the Rotor-Gene 6000 instrument (Corbett Life Science, Queensland, Australia). Each 10μL reaction contained 5μL Platinum SYBR Green qPCR SuperMix-UDG (2x, Invitrogen Life Technologies, New South Wales, Australia), final concentration of 0.5μM for each primer, 3.5μL ddH20, and about 20ng DNA template. The 10μL SYTO9 reaction included: 2.5μM SYTO9, final concentration of 0.5μM for each primer, 1.5mM $MgCl_2$, 0.2mM dNTPs, 1U platinum *Taq* DNA polymerase, PCR buffer (10x), about 20ng DNA template and ddH_2O to make up the remainder. The real-time PCR thermocycling parameters were: 50°C for 2min, 95°C for 2min, 40 cycles of (95°C for 5s, 60°C for 30s), 72 °C for 2min, and 50°C for 20s, followed by HRM ramping from 75–87°C with fluorescence data acquisition at 0.05°C increments. Reactions were routinely carried out in duplicate.

5.2.6   HRM Data Analysis

The Rotor-Gene 6000 proprietary software (version 1.7.34) enables the user to visualise HRM data in multiple ways. The negative derivative of a fluorescence (F) over temperature (T) (*df/dt*) curve displays the T_m, the normalised raw curve depicts the decreasing fluorescence vs increasing temperature, and difference curves (Wittwer *et al.*, 2003), that display a user-defined curve as the baseline (i.e., the x-axis), and depicts other normalised curves in relation to that baseline.

134

5.2.7 Criterion for Differentiating DNA Melting Curves

Criteria for calling melting curves as "same" or "different" using difference graphs were developed empirically. Melting curves are called the same as the defined control if the difference graph lies within ±4 U relative to the x-axis, and does not display reproducible differences such as double peaks or crossing the x-axis more than twice in both replicates. In addition, a 2-step procedure was followed to determine if an unknown HRM curve was the same as a known curve. First, the normalised HRM curve for the unknown type was compared to known normalised HRM profiles. These profiles are either generated together with the unknown samples, or more practically, have been produced previously. Second, the closest known HRM profile was selected as the difference graph control, and comparison of the difference curves was used to determine whether the unknown isolate was the same or different from the known. For each data analysis, the digital filter was set to "heavy" and the replicate grouping option was selected. When conflicts between replicates occurred, a repeat experiment was performed.

5.2.8 Inter-Run Comparison of High Resolution Melt Curves

The evaluation of inter-run variation was achieved by exporting the normalised data from separate experimental runs into Microsoft Excel. The quantitative inter-run variation was determined by manipulating the normalised HRM data to generate difference curves for each genotype. This was achieved by subtracting the normalised fluorescence values for each temperature point of a control genotype from the normalised data for each isolate. This data was imported into Teechart office where inter-run comparison of actual melting curve difference graphs could be visualised.

5.3 RESULTS

5.3.1 *spa Sequence Typing of Diverse Australian MRSA*

In order to select a diverse set of *spa* types to assess the HRM resolving power adequately, 34 MRSA isolates from 13 STs were subjected to *spa* DNA sequence analysis (Table 5.1). Sequencing revealed 22 *spa* genotypes with repeat regions containing from five (t631) to 12 (t018) 24bp repeats. Careful sequence analysis identified five novel types; three with novel combinations of known repeat units and two featuring new repeat units. Interestingly, the novel nine-repeat combination of isolate RPH29 (ST1) consisted of the first four repeats of t002 (ST5), repeat allele 34, and the last four repeats of t127 (ST1). This chimera arrangement provides evidence for *spa* homologous recombination between ST1 and ST5 strains.

5.3.2 *spa Sequence Variation Within Defined Genotypes*

The plan of selecting multiple isolates from identical STs was successful as many closely related *spa* types were identified. In total, 11 ST239 isolates yielded four related *spa* types, eight ST1 isolates produced three types; however the five ST30 isolates proved identical. With the exception of ST93, each remaining MLST/SNP genotype that contained more than a single isolate produced multiple *spa* subtypes. From the limited PFGE data for our collection, *spa* genotyping was able to subtype the two D pulsotype isolates, and notably, the ST239 isolates from six pulsotypes were divided into four *spa* types. Overall, with the identification of significantly different, closely related and identical *spa* sequence types, it was concluded that a robust assessment of *spa* HRM resolving power could be achieved.

Table 5.1. *spa* sequencing results

Isolates (*n*=44)	MLST[a] (*n*=13)	*spa* type [b] (*n*=22)	*spa* repeats [c]
SN39	1	t127	07-23-21-16-34-33-13
FMC53	1	t127	07-23-21-16-34-33-13
SJOG 30	1	t127	07-23-21-16-34-33-13
MD828534	1	t127	07-23-21-16-34-33-13
RHH10	~1	t127	07-23-21-16-34-33-13
FH53	~1	t127	07-23-21-16-34-33-13
749	1	t175	07-23-21-16-16-33-21-16-33-13
RPH29	~1	NEW1	26-23-17-34-34-16-34-33-13
E804531	5	t002	26-23-17-34-17-20-17-12-17-16
37	~5	t002	26-23-17-34-17-20-17-12-17-16
CH 97	73	t002	26-23-17-34-17-20-17-12-17-16
24	~5	NEW2	<u>na</u>-23-17-34-17-20-17-12-17-16
56	~5	NEW2	<u>na</u>-23-17-34-17-20-17-12-17-16
IMVS 67	8	t008	11-19-12-21-17-34-24-34-22-25
251	8	t008	11-19-12-21-17-34-24-34-22-25
RPH 2	8	t190	11-17-34-24-34-22-25
CH69	~22	NEW3	26-23-13-17-31-29-17-25-17-25-16-28
68284/98	30	t019	08-16-02-16-02-25-17-24
62 (1)	30	t019	08-16-02-16-02-25-17-24
PAH 1	30	t019	08-16-02-16-02-25-17-24
PAH 58	30	t019	08-16-02-16-02-25-17-24
D828570	~30	t019	08-16-02-16-02-25-17-24
E822485	36	t018	15-12-16-02-16-02-25-17-24-24-24
B8-31	~45	t123	09-02-16-34-13-16-34-16-34
MJ710566	~45	t065	09-02-16-34-13-17-34-16-34
337	59	t437	04-20-17-20-17-25-34
751	59	t216	04-20-17-20-17-31-16-34
IP01M1081	59	t216	04-20-17-20-17-31-16-34
IP01M2046	78	NEW4	08-21-17-13-13-<u>na</u>-34-33-34
MC801535	88	t325	07-12-21-17-34-13-34-34-33-34
F829549	88	t186	07-12-21-17-13-13-34-34-33-34
RBH 98	93	t202	11-17-23-17-17-16-16-25
15808-0020	~93	t202	11-17-23-17-17-16-16-25
15814-9852	~239	t631	15-12-16-17-24
RPAH 18	239	t037	15-12-16-02-25-17-24
AH 13	239	t037	15-12-16-02-25-17-24
IMVS 20	239	t037	15-12-16-02-25-17-24
K714372	~239	t037	15-12-16-02-25-17-24
K711532	~239	t037	15-12-16-02-25-17-24
K704540	~239	t037	15-12-16-02-25-17-24
E812560	~239	t1155	15-12-16-02-25-17-16
IPOOM14235	~239	t1155	15-12-16-02-25-17-16
MK703484	~239	t1155	15-12-16-02-25-17-16
14176-5710	~239	NEW5	15-21-12-16-02-25-17-16

[a] "~" denotes an isolate genotyped using the high *D* SNP set.

[b] NEW = New *spa* type

[c] <u>na</u> = New allele

Blind analysis isolates in bold.

5.3.3   HRM Performance Comparison of SYBR Green and SYTO9

Real-time PCR requires a fluorescent dye to detect increasing double-stranded DNA levels, whereas HRM analysis detects a reduction in double-stranded DNA levels. As a sensitive and stable fluorescent dye is required for these tasks, the commonly used SYBR green dye was compared with the newer SYTO9 dye to determine which produced the best results. The major functional difference between the two is that SYBR green molecules intercalate irregularly into amplified double stranded DNA, whereas the SYTO9 intercalates into each DNA groove. This has been suggested to allow greater sensitivity and to provide more accurate results. Fig 5.1 shows the HRM comparison and shows that SYBR green performed better. Therefore, it was decide to use SYBR green exclusively for the remainder of the study.
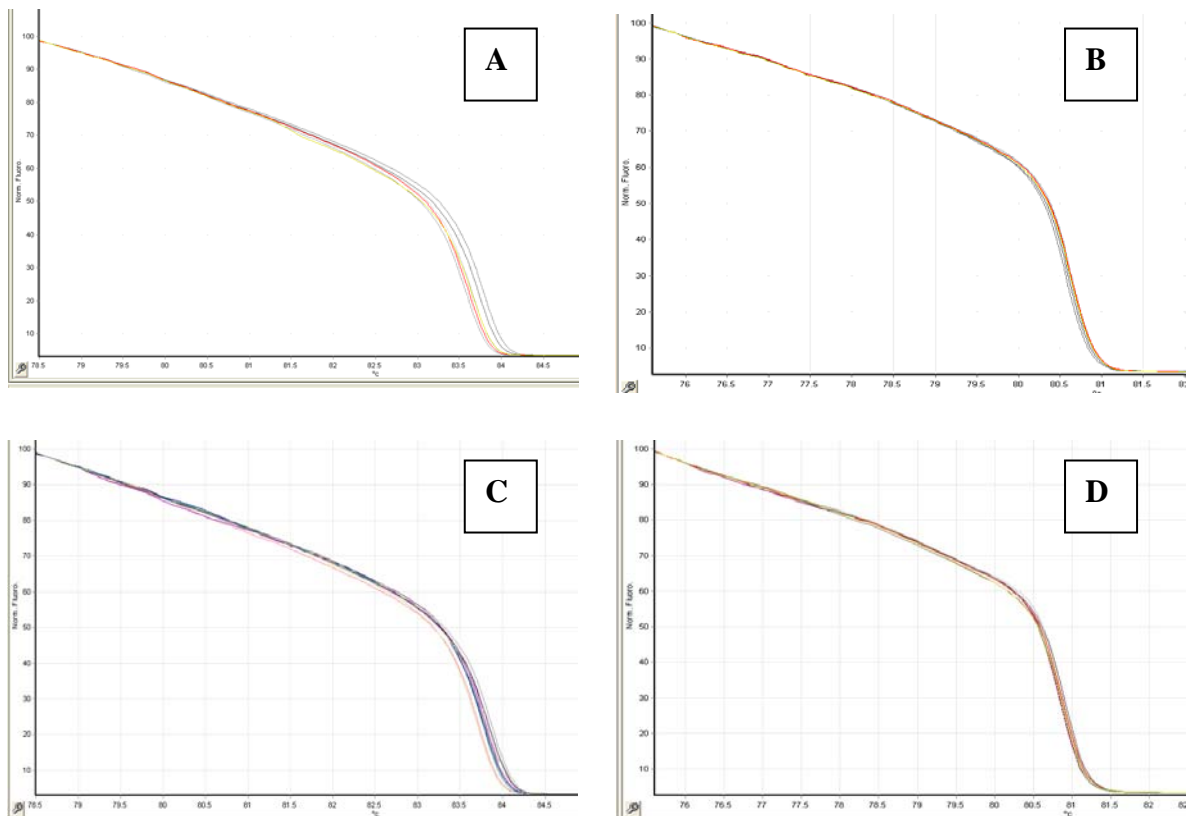
Fig 5.1: Comparison between SYBR Green and SYTO9 HRM performance. From these normalised HRM curves it can be seen that the SYBR green (B and D) displayed greater reproducibility than SYTO9 (A and C). Pooled ST30 isolates (A and B), and pooled ST239 isolates (C and D) were used for the comparison.

5.3.4   HRM Curves from Hypervariable spa Sequence

The HRM methodology was initially tested against 22 known *spa* sequences that were derived from 34 diverse MRSA isolates during the course of this study (Table 5.1). According to the classification criteria stated (Fig 5.2), the 22 known *spa* sequences generated 20 different HRM curves (Table 5.2). A difference graph comparing the melting curves produced from the four *spa* types carrying eight repeat units is show in Fig 5.3. Two pairs of *spa* sequences were not resolved by HRM analysis; these were t002 and NEW2, and t037 and t1155. Both pairs of sequences differed at a single nucleotide. We concluded that *spa* sequences can be differentiated by HRM even when they are closely related, but the resolving power using SYBR Green dye does not extend reliably to sequences that differ by a single base pair.
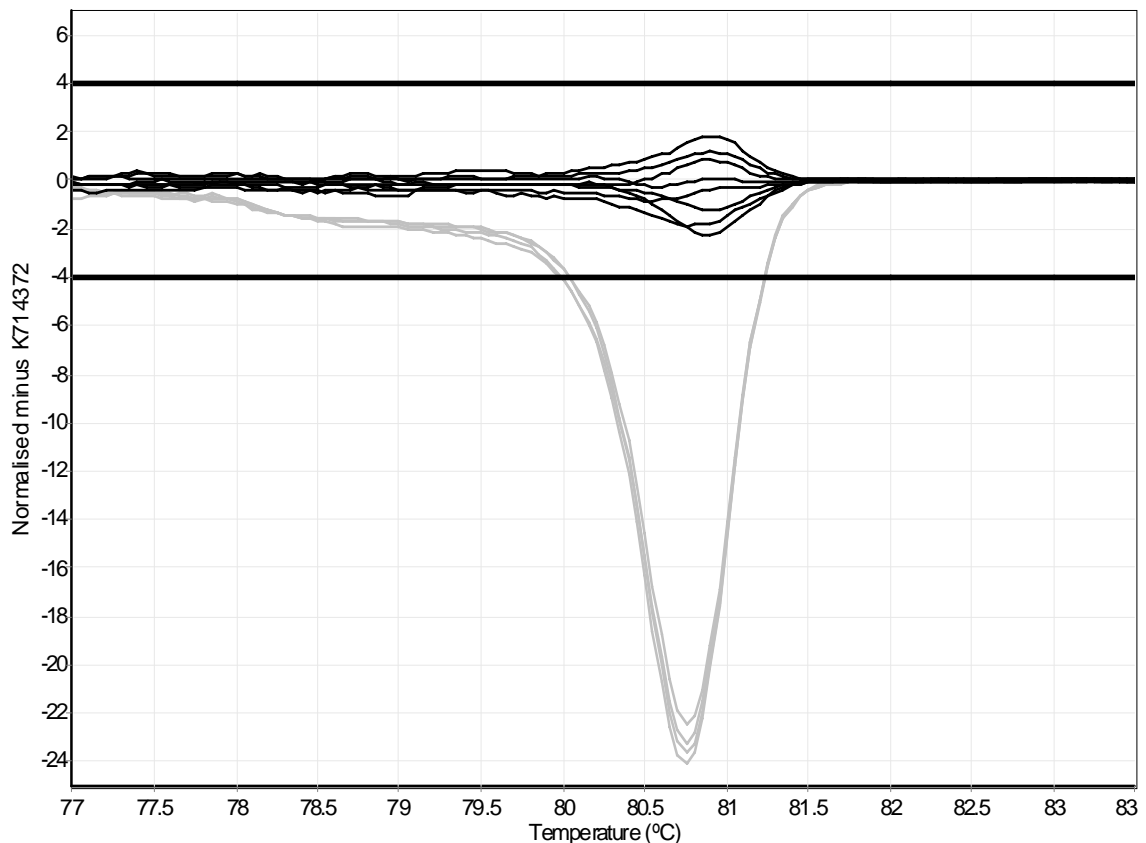
Fig 5.2. A typical difference graph comparison of two unrelated *spa* HRM genotypes. From the six isolates displayed, four fall within the ±4 units of the baseline profile (HRM 19) and are called "same", the remaining 2 isolates lie outside of the ±4 cut-offs and are therefore denoted as "different" (HRM 1). In this instance, to highlight the HRM reproducibility within a run, the replicates were not grouped together.

139

Table 5.2. *spa* **HRM genotypes**

Isolate	MLST[a] (*n*=13)	*spa* genotype[b] (*n*=22)	HRM genotype (*n*=20)	*spa* repeat units[c]	T_m[d]	Repeat region GC%	Repeat region size (bp)
SN39	1	t127	HRM1	07-23-21-16-34-33-13	80.6	44	168
749	1	t175	HRM2	07-23-21-16-16-33-21-16-33-13	82.1	46.3	240
RPH29	~1	**NEW1**	HRM3	26-23-17-34-34-16-34-33-13	79.9	41.7	216
E804531	5	t002	HRM4	26-23-17-34-17-20-17-12-17-16	81.6	45.4	240
24	~5	**NEW2**	HRM4	**new**-23-17-34-17-20-17-12-17-16	81.6	45	240
IMVS 67	8	t008	HRM5	11-19-12-21-17-34-24-34-22-25	80.6	43.8	240
RPH 2	8	t190	HRM6	11-17-34-24-34-22-25	80.0	44	168
CH69	~22	**NEW3**	HRM7	26-23-13-17-31-29-17-25-17-25-16-28	81.4	44.4	288
68284/98	30	t019	HRM8	08-16-02-16-02-25-17-24	80.7	44.3	192
E822485	36	t018	HRM9	15-12-16-02-16-02-25-17-24-24-24	81.3	44.7	264
B8-31	~45	t123	HRM10	09-02-16-34-13-16-34-16-34	79.6	40.7	216
MJ710566	~45	t065	HRM11	09-02-16-34-13-17-34-16-34	79.7	40.7	216
337	59	t437	HRM12	04-20-17-20-17-25-34	80.4	43.5	168
751	59	t216	HRM13	04-20-17-20-17-31-16-34	81.0	44.3	192
IP01M2046	78	**NEW4**	HRM14	08-21-17-13-13-**new**-34-33-34	79.7	40.7	216
MC801535	88	t325	HRM15	07-12-21-17-34-13-34-34-33-34	79.7	40.4	240
F829549	88	t186	HRM16	07-12-21-17-13-13-34-34-33-34	79.8	40.8	240
RBH 98	93	t202	HRM17	11-17-23-17-17-16-16-25	82.2	47.9	192
15814-9852	~239	t631	HRM18	15-12-16-17-24	81.1	47.5	120
RPAH 18	239	t037	HRM19	15-12-16-02-25-17-24	80.9	45.2	168
E812560	~239	t1155	HRM19	15-12-16-02-25-17-16	80.9	45.2	168
14176-5710	~239	**NEW5**	HRM20	15-21-12-16-02-25-17-16	81.2	45.3	192

[a] "~" denotes isolates genotyped using MLST SNP typing.

[b] **NEW** = new *spa* type.

[c] **new** = new repeat unit.

[d] T_m calculated by the Rotor-Gene 6000 using SYBR Green dye and rounded to one decimal place.
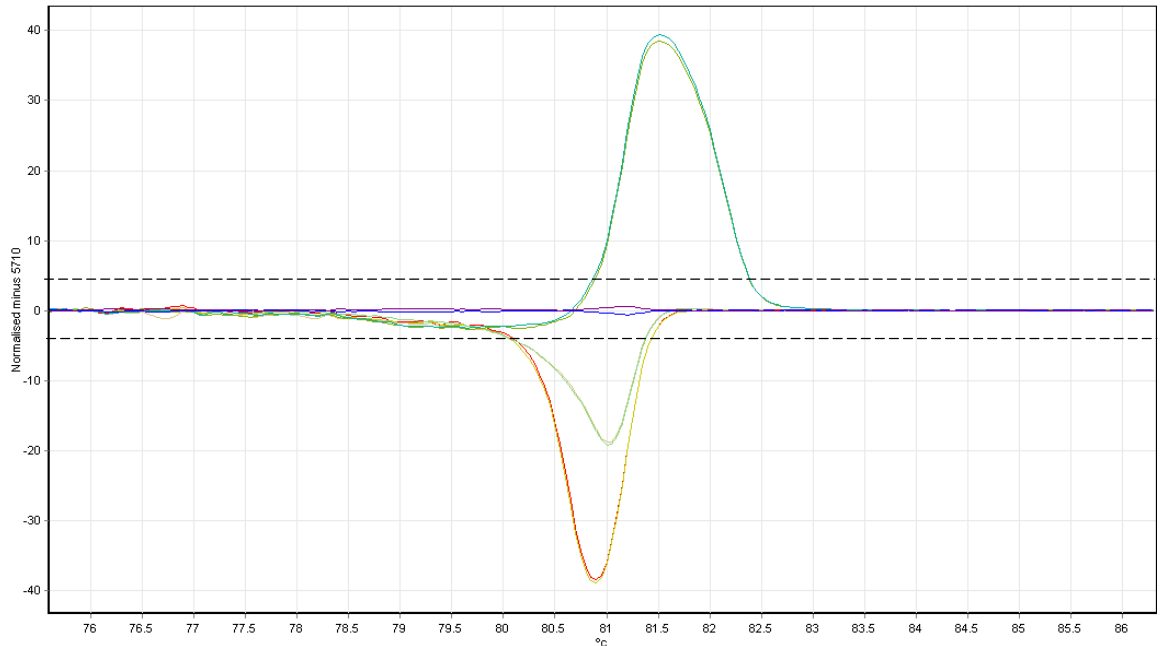
Fig 5.3. Rotor-Gene software difference graph showing the disparity between the *spa* **types carrying eight repeat units.** HRM profiles from top to bottom; HRM17, HRM20 (baseline), HRM13 and HRM8 (Table 5.2). Despite each isolate containing an identical number of repeat units, the difference graph demonstrates that the HRM protocol can easily differentiate each melting curve. This also demonstrates the large contribution of GC% to HRM melt curve differences. For this graph the "group replicates" option was not selected and the dotted lines were manually added to depict the ±4 thresholds that specify whether a curve is called the "same" or "different".

5.3.5   HRM Resolution Compared to T_m

Comparison between the T_m and the HRM profiles revealed that the HRM results showed better resolution compared to that of T_m alone (Table 5.2, Fig 5.4). For example, HRM curves 1 and 5 were derived from different *spa* VNTR sequences but have identical T_m values, as was the case with HRM curves 11, 14, and 15. Although the *df*/*dt* curves resolve the majority of HRM *spa* profiles, it is not possible to develop a practical genotyping criterion as there is no data normalisation function. Furthermore, since the curves are very similar in shape, it is difficult to differentiate closely related strains when subtle experimental variation is considered. The HRM differences between three similar *df*/*dt* curves are shown in Fig 5.5.
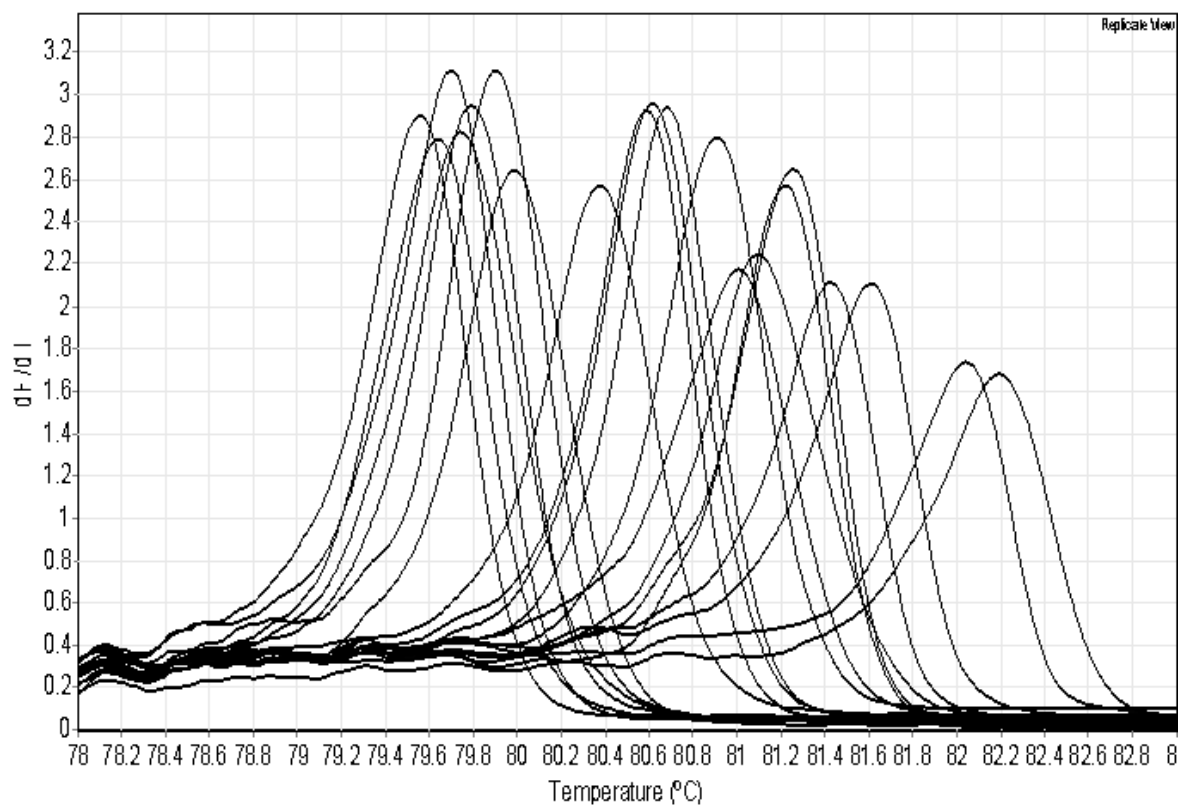
Fig 5.4. Representative *df/dt* curves indicating the T_m range for the 20 HRM profiles.
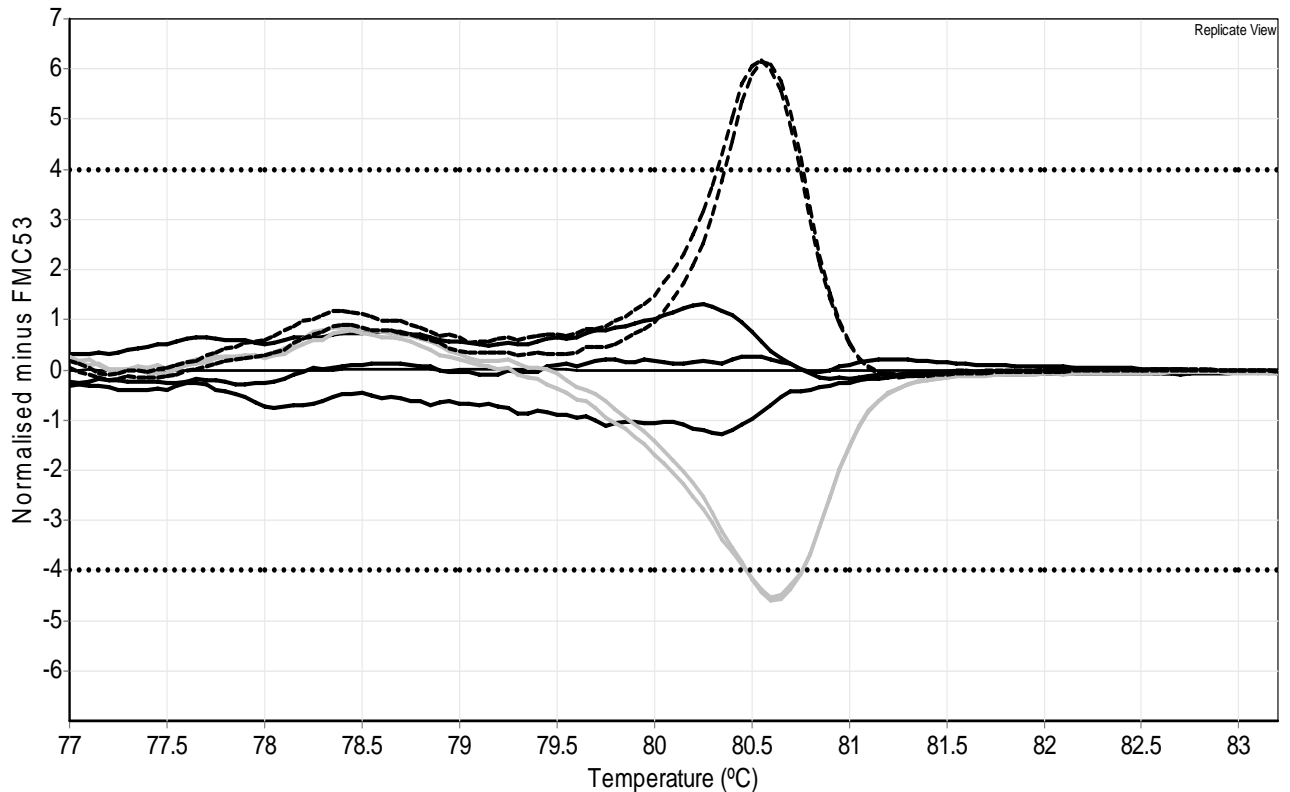
Fig 5.5. Differentiation of three similar *spa* HRM curves using difference graphs. The three displayed *spa* HRM profiles correspond to the three *df/dt* curves located at 80.6°C on Fig 5.4. The genotypes from three isolate are identical to the baseline *spa* profile (HRM 1); two isolates rise above +4 (HRM 8), two isolates fall beneath –4 (HRM 5) and are each denoted as "different." The Y-axis label indicates the isolate the unknown *spa* HRM types were normalised against.

5.3.6   Blind Analysis Testing of the HRM Reproducibility

To confirm the robustness of this method, two blind experiments were performed with different batches of reagents used each time. The first experiment entailed the analysis of 10 MRSA isolates of unknown *spa* sequence, as indicated in Table 5.1. s*pa* sequence determination revealed that all the *spa* types were identical to existing types in the collection, and all were identified correctly by HRM analysis. The second blind experiment involved reanalysis of the original set of 34 isolates. Once again, the *spa* types were 100% consistent with the first two times these isolates were analysed. Interestingly, because the melt curves were clearly visually distinct, the *spa* sequences differing by only a single base could be differentiated even though they

143

fell within the conservative sequence discrimination criteria. It is likely that these criteria will be refined as more data become available, thus increasing the resolving power of HRM analysis.

5.3.7   Inter-Run Comparison of Melting Curves

The software supplied with the Rotor-Gene 6000 does not allow generation of difference graphs using curves from different runs of the instrument. To determine the practicality of HRM curve comparison among runs (and by extension, the development of a standard library of HRM curves), normalised data were exported and converted to "difference" format in Microsoft Excel, and difference curves assembled using the freeware chart drawing program Teechart Office. The HRM curves were completely portable and the practice of comparing data from different real-time PCR runs caused no loss in the ability to differentiate different *spa* alleles Fig 5.6.
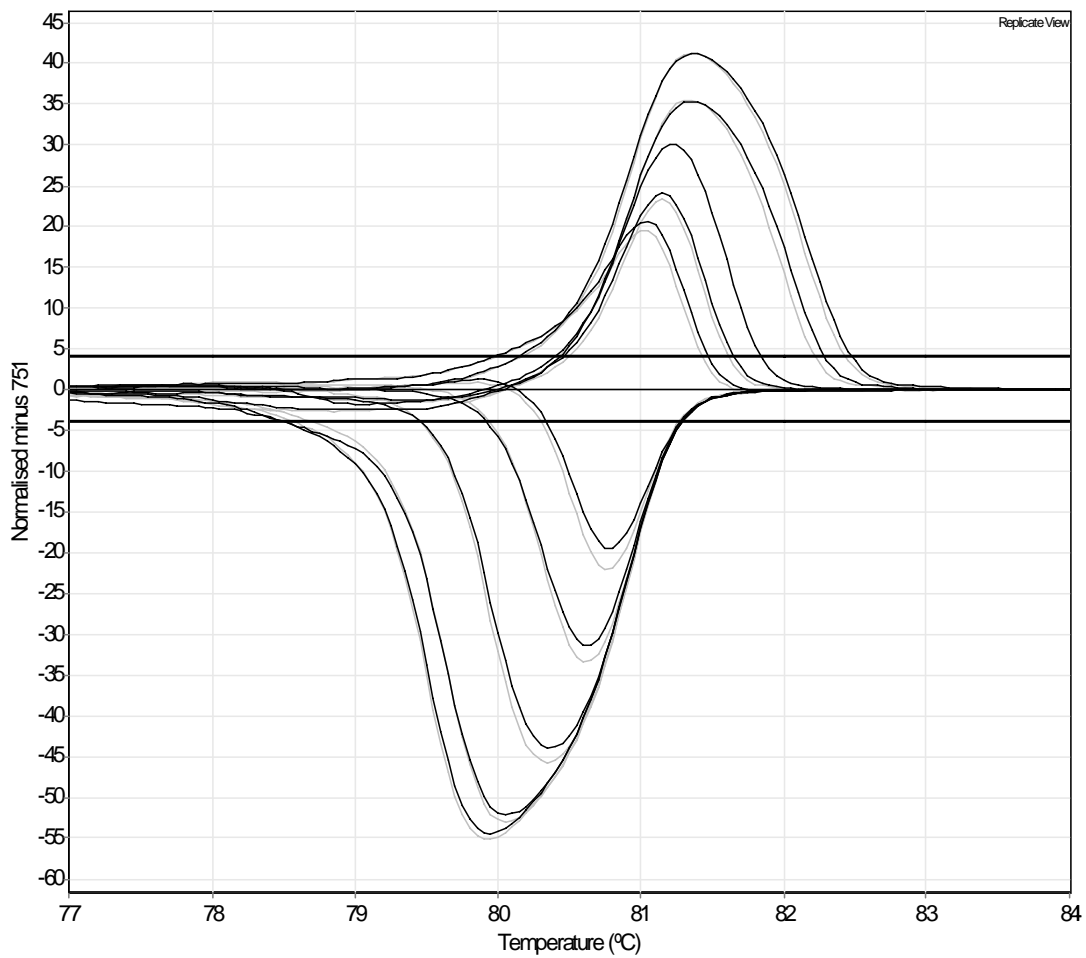
Fig 5.6. Difference graph demonstrating the accurate reproduction of 11 *spa* **HRM profiles from separate experiments.** The comparison of 11 *spa* HRM profiles from two runs (grey and black traces) demonstrate that *spa* HRM analysis is potentially portable in its results output.

145

5.4 DISCUSSION

Rapid genotyping of pathogenic bacterial strains is important for infection management and control of disease transmission. The *spa* VNTR region of *S. aureus* undergoes a high rate of mutation that enables genotyping of closely related strains by DNA sequencing (Shopsin *et al.*, 1999). The *spa* sequence types can either be added to MLST to generate highly informative combinatorial genotypes, or simply be used in isolation. The widespread uptake of the method is demonstrated by the large online database that catalogues extensive known diversity. The cost, experimental effort and time taken to achieve a sequencing result, however, is not desirable for rapid diagnostic methodologies. Therefore, the chief aim of this study was to develop a *spa* VNTR genotyping method that did not require DNA sequencing.

Rapid cycle real-time PCR followed by HRM analysis has emerged as a practical and effective method for detecting DNA sequence variation. The central principle of HRM analysis is that the dissociation dynamics of double-stranded DNA depends on nucleotide composition, in addition to T_m. As the G:C binding affinity is stronger than A:T, the distribution of each nucleotide pair across a heated DNA fragment influences the dissociation profile. HRM analysis has advantages over standard T_m measurement due to a proven ability to differentiate fragments that display similar lengths and T_ms. This is because measuring the real-time dissociation progress of double-stranded DNA accommodates subtle binding variations that result from each nucleotide being influenced by its proximal nucleotides (Breslauer *et al.*, 1986).

Here, SYBR green real-time PCR and HRM analysis was applied to 22 *spa* sequence types, and revealed that the methodology is sufficiently sensitive to produce 20 discrete HRM curves. VNTR regions from the 22 *spa* sequences ranged in length from 120-288bp and displayed GC% from 40.4-47.9%. Judging from the sequence characteristics of the shortest and longest fragments, it appears that the GC% can influence the T_m value more than sequence length. Due to the small fragment length and composition of repeating units, each fragment dissociated uniformly and displayed a clean single peak on the *df/dt* graph. Double or triple peaks can arise through staggered dissociation patterns resulting from GC% discrepancies between different regions of the fragment, and also if multiple amplicons are present. For this

146

reason, multiplexing the *spa* PCR with other reactions is not possible unless amplicon specific dyes are incorporated in the experiment.

The use of raw normalised and difference graphs allowed whole curve analysis and the opportunity to create a criterion for calling curves that were same or different. In most cases, the raw normalised curves were sufficiently different to imply variation in the amplified sequence. For closely related *spa* types visualised on the raw normalised curves, reproducible differences were often subtle and required the difference graphs for genotype assignment. The cut-off of ±4 units in relation to the x-axis using the difference graph was formulated by comparing the experimental variation resulting from identical known isolates over multiple runs. This value was set conservatively as HRM profiles for *spa* types differing by a SNP actually displayed minute, but reproducible differences. In the small number of instances where an unknown *spa* HRM profile appeared borderline, it was useful to consider the shape of the difference curve in relation to other isolates identical to the control. Identical isolates typically did not cross the x-axis, whereas borderline curves often displayed reproducible differences including crossing the x-axis multiple times or generally displaying a subtly different shape. In such instances it was also useful to refer back to the raw normalised and conventional *df/dt* graphs to assist with the classification.

Initially, it was only possible to compare *spa* HRM curves accurately that were generated in the same experiment. Inter-run comparison of *spa* HRM curves was only possible by opening the runs in separate copies of the Rotor-Gene software, and then relying on visual comparison while switching between the programs. In order to facilitate an actual comparison of inter-run data, a method was developed based on exporting and manipulating the raw data into Microsoft Excel, and then comparing the raw normalised or difference graphs in the graphing program, Teechart office. This process effectively allowed HRM curves to become portable, which is very significant as this characteristic makes the method potentially library based rather than strictly comparative. One resource that would be very useful in the future is a library of *spa* HRM curves for the major MRSA epidemic and community-acquired clones, and this library has been provided, in part, by this study. However, for this to be achieved effectively, strict standardisation measures would be required and may involve the testing of control isolates.

It was a simple matter to assign HRM profiles to known genotypes, therefore, we utilised two blind analyses to assess the sensitivity and reproducibility of the assay and also the performance or our classification criteria. The first blind analysis consisted of generating HRM profiles from 10 MRSA isolates of unknown *spa* types, and then comparing them to known HRM curves using the inter-run comparison method. The curves from the 10 isolates were found to match three previously resolved profiles that sequence analysis confirmed. Reproducibility and robustness of the assay and our criteria was provided additional confirmation by successful blind HRM genotyping of the original 34 isolates. However, considering the numerous *spa* sequence types identified worldwide, it is likely that HRM analysis of larger collections will result in increased instances of identical melt curves for different sequences. This can be overcome easily by coupling *spa* VNTR HRM with a clonal genotyping method.

Overall, including the optimisation phase, the HRM analysis was performed four times on the 34 isolates, and the results proved to be completely reproducible with respect to defining the *spa* types. However, there was one very minor reproducibility issue. When the Corbett Rotor-Gene 6000 runs were separated by several weeks, the HRM curves from the final experiment were found to be offset by approximately 0.2°C, although there was complete consistency of curve shape. These results were likely attributable to very slight changes to the calibration of the instrument's thermometer and can easily be overcome by including a known control in all runs.

Since real-time PCR analysis has become common place, reagent companies have produced fluorescent dyes that are designed to provide better performance. Recent real-time PCR dyes are designed to intercalate into each groove of the double stranded DNA so greater resolution can be achieved for HRM applications that analyse single polymorphisms. However, results from the current study have determined that the older SYBR green actually performed better in terms of HRM curve consistency, even though it intercalates less frequently. Furthermore, as SYBR green is combined with a polymerase, buffer, $MgCl_2$ and dNTPs in a master mix, it is very convenient.

In conclusion, HRM analysis of the *spa* VNTR locus is an effective method for the easy and rapid identification of *spa* alleles. The major advantage of this method is that it is a single-step and closed-tube procedure performed on a moderately priced and generic piece of laboratory equipment. This technique offsets the occasional inability to differentiate *spa* alleles, because *spa* interrogation could simply be combined with the real-time PCR interrogation of, for example, clonal-complex–specific single-nucleotide polymorphisms, toxin-encoding genes, or binary markers that subtype SCC*mec*.

CHAPTER SIX

GENERAL DISCUSSION

6.1 DISCUSSION

Suitable bacterial genotyping methods are essential for the management of infectious diseases in both healthcare and community environments. Subspecies identification is conducted within hospitals and research institutions for many applications, including for identifying an outbreak source during epidemiological investigations, for routine surveillance of hyper-invasive, multi-resistance or simply carriage strains, and for small or large scale population structure analyses. Infectious disease genotyping is also utilised for biodefence in order to identify the source of biological agents.

The bacterial genotyping methods currently available vary in their resolving power, cost effectiveness, and the technological platform required to employ them. Based on these factors, the most appropriate genotyping method for any given application can be determined. Studies involved in detailed investigations of virulence or other complex traits require whole genome sequencing or microarray analysis. Both approaches are time consuming and expensive. In contrast, routine analysis of clinical isolates requires genotyping methods that are rapid, cost effective, and generate portable results of appropriate resolution that can be interpreted easily. This thesis responds to a requirement for such a method through development of a comprehensive *S. aureus* genotyping strategy underpinned by three novel methods based upon the PHRANA model (Keim *et al.*, 2004).

Until recently, molecular genotyping of MRSA was largely achieved by PFGE of macro restricted genome fragments. PFGE produces a high level of genotyping resolution resulting from restriction site polymorphisms and fragment length variation. Accordingly, PFGE has contributed significantly to the understanding of MRSA transmission on a local and global scale (Oliveira *et al.*, 2001)**.** Several limitations of the PFGE method exist, however, and these relate to inter-laboratory reproducibility, result interpretation, and experimental turn-around-time, which have made modern comparative genomic methods that produce genetic fingerprints based on known differences more appealing. In due course, it is likely that PFGE will be superseded for the routine genotyping of *S. aureus*.

Numerous bacterial genotyping methods have been developed since PFGE, and of them, MLST is probably the most significant. The benefit of MLST when compared with PFGE is the digital nature of the data, which can portray unambiguously the differences between strains in relation to their population structure. DNA sequence-based genotyping methods produce portable results that can be stored and accessed readily via the internet, thereby providing a comprehensive record of global diversity for the target bacteria. MLST is suited ideally for long-term analysis of *S. aureus* as it surveys slowly accumulating SNPs within the conserved housekeeping genes of the core genome. A major disadvantage of MLST is its reliance on the generation of seven ~450bp nucleotide sequences. Even with current sequencing technology, this can cause problems when analysing large numbers of isolates in a routine clinical microbiology laboratory. A further shortcoming of MLST is its lack of resolution for short-term investigations or outbreak situations where a high degree of genotyping power is required. In these circumstances, MLST resolving power is often improved by addition of complementary genotyping methods that target regions inherently more variable than MLST polymorphisms.

SCC*mec* characterisation and MLST in combination provides two main benefits: SCC*mec* structural variation increases genotyping power and, more importantly, provides a method for tracking difficult-to-treat MRSA strains. The SCC*mec* elements belong to a family of heterogeneous MGEs that are responsible for transmission and expression of β-lactam resistance in MRSA. Inherent SCC*mec* variation, combined with emergence of CA-MRSA, has fuelled development of numerous SCC*mec* genotyping methods. Each of the current methods provides a compromise between the number/type of targets chosen and the resolution achieved. The two commonly used methods are the simplex and the multiplex methods, which target the *ccr/mec* complex types and SCC*mec* type-specific domains, respectively. Although a SCC*mec* genotyping method with better resolution has been developed, it requires analysis of 27 PCRs within six multiplex reactions (Kondo *et al.*, 2007). Thus, given the extensive knowledge of SCC*mec* variation and the demand for economical and high throughput clinical analysis, the need for improved genotyping power has remained.

The legacy of PFGE as the premier *S. aureus* genotyping method has set a high resolution precedent; and as a consequence, analysis of the hypervariable *spa* VNTR

154

locus has become a well established method. The utility of *spa* VNTR sequencing as a single genotyping strategy is justified by the automated and objective algorithm known as Based Upon Repeat Pattern (BURP) (Strommenger *et al.*, 2006; Mellmann *et al.*, 2007). BURP analysis identifies *spa* CCs based on relationships among *spa* sequence types. After comparison with CCs defined by MLST and microarray analysis, the BURP *spa* CCs displayed comparable results (Mellmann *et al.*, 2007). The powerful ability of the *spa* VNTR to subtype MLST and PFGE types, however, is largely ignored by BURP analysis.

There is currently no standardised combined genotyping strategy that includes the *spa* VNTR region. *S. aureus* epidemiological investigations have used *spa* sequence typing with MLST (Layer *et al.*, 2006), while other studies have recommended *spa* typing as an ideal partner with PFGE (Faria *et al.*, 2008). The *spa* VNTR variation is also featured in the powerful *S. aureus* Multiple-locus VNTR Analysis (MLVA) method (Sabat *et al.*, 2003), which discriminates genotypes based on multiple VNTR PCR fragment sizes (Keim *et al.*, 2000). Although there is justification for using *spa* VNTR sequencing as an independent genotyping method, there is evidence suggesting that the *spa* VNTR is most appropriately used in combinatorial genotyping approaches that match the genotyping power of PFGE.

The combination of MLST, SCC*mec* and *spa* typing conveniently embodies the benefits of the PHRANA concept, as a genotyping resolution comparable to PFGE is produced. A study from Denmark exemplified the power of this combined strategy with seven ST5 isolates showing three SCC*mec* types but with an identical *spa* sequence type, 72 ST8 isolates producing one SCC*mec* type and five *spa* sequence types and six ST45 isolates producing two SCC*mec* types and five *spa* sequence types, respectively (Bartels *et al.*, 2007). Even though this strategy produced MLST subtypes and defined the MRSA lineage, the methods are technically demanding and therefore are not used commonly together. In many cases, large scale MRSA studies use PFGE at the initial screening stage, then genotype a representative set of isolates using all or combinations of the three aforementioned methods. Taken together, it is clear that although MLST, SCC*mec* and *spa* typing is a very useful PHRANA MRSA genotyping strategy, technical difficulties and high costs associated with large scale or routine use of these methods have meant PFGE remains the popular choice.

The central aim of the current study was to develop alternative approaches to MLST, SCC*mec* and *spa* typing for the purpose of creating a cost and time efficient combined genotyping strategy for MRSA. The results identified a small number of highly informative genotyping targets that can be interrogated easily on a single and routine experimental platform. The strategy was based on the PHRANA model that enables high genotyping resolution by interrogation of low resolution markers in combination with markers exhibiting higher resolution as a result of genetic instability (Keim *et al.*, 2004). PHRANA combined genotypes are highly informative as they in the first instance divide species into broad ancestral clones, and then subdivide clones into strains and sub-strains. The overall genotypes are equally adept at placing unknown isolates from both long-term and short-term investigations accurately into the population structure. While no formalised three tiered PHRANA MRSA genotyping methodology exists, currently, the use of MLST, SCC*mec* and *spa* typing is an excellent example of such a strategy.

Development of a potentially routine MRSA PHRANA genotyping strategy required three separate and economical genotyping methods that targeted gene regions with varying evolutionary rates. Chapter Three of this thesis described the foundational method for the genotyping strategy, which is based on a set of highly informative SNPs derived from the MLST database. SNPs were identified using Minimum SNPs and classified *S. aureus* into broad groups of CCs defined by their genetic background. The high *D* SNP set were interrogated using allele specific real-time PCR assays and were tested on a diverse collection of MRSA. This set performed well for genotyping both known and unknown isolates into correct MLST defined CCs. The high *D* SNP set is able to generate meaningful groups efficiently that correlate with the MLST population structure because the *S. aureus* recombination rate is low. The only disadvantage of the SNP typing method is the occasional dissimilarity between ST and SNP group, a problem that could be alleviated by addition of a small set of binary virulence genes and antimicrobial resistant plasmids. Overall, the high *D* SNP set produced genotyping results that were largely congruent with the established MLST population structure and provided a suitable base method for a combined genotyping strategy.

The second component of the genotyping strategy results from systematic identification of highly informative SCC*mec* binary markers using Minimum SNPs. We hypothesised that when used in combination with the MLST SNPs they would increase the resolving power whilst simultaneously defining the MRSA lineage. The power of Minimum SNPs has been demonstrated previously for the identification of highly informative genotyping SNPs from MLST databases; however, results presented in Chapter Four demonstrates the capability of the software to identify sets of high performance genotyping targets from a data set reflecting binary marker variation, rather than SNP allelic variation. In order to identify targets that genotype each major and minor SCC*mec* type, a comparative data set was constructed of 46 SCC*mec* variants, each converted into a binary code based on the presence or absence of 34 genetic markers known to compose the SCC*mec* elements. Rather than identifying single points of variation that relate specifically to a single genotype, the Minimum SNPs *D* algorithm considers each point of variation systematically within the input data set and calculates genes that in combination generate optimal genotyping power. In comparison with the standard targets used for SCC*mec* characterisation, the Minimum SNPs binary markers displayed higher target versus type ratios, which demonstrates the effectiveness of the computerised approach. Although the presence/absence testing of the SCC*mec* targets was not reduced to experimental practice, potential users instead have a choice between target sets that optimally meet their genotyping needs.

The high *D* SNP set, in combination with the SCC*mec* binary targets, is a practical alternative to the current methods that target corresponding loci. To generate a level of resolution equivalent to that of PFGE, however, interrogation of another highly variable region is required. Chapter Five describes the final component of the PHRANA strategy, which uses real-time PCR and HRM to analyse the hypervariable *spa* VNTR region. HRM analysis has the potential to act as an alternative to DNA sequencing since it can be used to differentiate fragments based on sequence abnormalities. When tested on a diverse subset of MRSA isolates from our collection, HRM analysis defined 20 of the 22 *spa* sequence types. When tested on STs that were grouped inappropriately using the high *D* SNP set, HRM analysis was able to differentiate each ST via differences in melting curves. In addition to the high resolution obtained using this method, HRM also exhibited high reproducibility

within and between experimental runs. This was verified twice in blind analyses. The ability to compare HRM profiles among runs was a significant result and allows for construction of a database of melt curves/data that can be compared with unknown isolates to determine type. It is feasible that *spa* VNTR HRM analysis could be used as a rapid stand-alone genotyping method. Overall, the performance of the *spa* VNTR HRM analysis cemented its place as the third tier in the real-time PCR MRSA genotyping strategy.

The combined use of the three aforementioned methods effectively brings the genotyping power of the MLST, SCC*mec* and *spa* loci to any routine clinical microbiology laboratory. The PHRANA strategy underpinned by Minimum SNP-derived targets have the potential to provide high quality outcomes that can be used for most genotyping applications. Since each strategy component utilises the real-time PCR approach, these methods are simple to prepare, low-cost and timely. Furthermore, it is also possible to employ the methods in alternative combinations. For example, if full MLST is a possibility, then the use of the SCC*mec* binary targets has the potential to greatly enhance MRSA characterisation compared with current methods. In the case of MSSA genotyping, just the high D SNPs and *spa* HRM provides rapid and adequate genotyping resolution. Finally, *spa* HRM alone exhibits resolution suitable for answering the question "are these isolates the same or different?" Overall, the ability to derive high resolution genotyping targets from regions of complex variation and the ability to characterise VNTR regions without the use of DNA sequencing are two significant contributions made here.

The genotyping efficacy of targets derived from dynamic comparative genomic databases can be affected by subsequent database expansion. Two considerations are pertinent, the extent of species variation within the initial data set and the mutation rate of the analysed region. Minimum SNPs input data sets should optimally reflect the diversity seen in nature as this is important for identifying genotyping targets that are broadly applicable. The MLST SNPs were selected from a data set that at the time featured the major clones and therefore were able to perform well on diverse sets of known and unknown isolates. As the MLST housekeeping genes mutate slowly, expansion of the MLST database generally results from the discovery of new STs that belong to existing clonal complexes rather than singletons. The expansion of clonal

complexes serves to decrease the overall *D* value of the SNP set, however it does not affect the robustness of the resolved groups, unlike singletons. Since no major clones have been added to the MLST database since the high *D* SNPs were identified, they have largely retained their genotyping power, as was demonstrated by Dr. Flavia Huygens who upon reanalysis of an updated data set identified largely similar targets (data not shown). This occurred because Minimum SNPs identifies SNPs that mutated early in the clonal expansion of *S. aureus*. Considering the ancient SNPs were identified from a data set representing adequate diversity and undergo a slow mutation rate, the high *D* SNPs are likely to hold their value and not be affected too greatly by homoplasy.

These considerations also refer to the SCC*mec* targets, but are likely to impact genotyping performance more than the MLST SNPs. Firstly, the data set of SCC*mec* variation was compiled during this study and did not result from a comprehensive global comparative database like that of MLST. This means that undiscovered SCC*mec* variants have the potential to affect the efficacy of the target sets identified in this study, as novel SCC*mec* variants could be incorrectly genotyped. This is more likely to occur, compared to the MLST SNPs, due to a significantly smaller number of polymorphic sites (34 binary markers compared to 3198 potential SNPs) and a faster SCC*mec* evolution rate. The SCC*mec* elements can significantly change states in short time due to spontaneous MGE insertion/deletion events that modify *mec* complexes and produce new junkyard configurations. These rapid evolutionary changes are somewhat counteracted, however, by the inclusion of *ccr* alleles as they undergo slower evolution of actual DNA sequence. It is foreseeable that many novel SCC*mec* variants will be discovered; therefore it is likely that the performance of the SCC*mec* genotyping target sets will be reduced. However, as the target sets produce predefined results, the effect of novel types can be determined and therefore, subsequent targets sets that accommodate novel variants can be easily calculated if required.

SCC*mec* binary diversity was used here to identify high performance genotyping targets. It is entirely feasible, however, to use this as a model to analyse other databases of complex diversity. The prime comparative methods, apart from MLST, that produce massive and compatible database variation are whole genome sequencing and microarrays.

Since the process of generating a whole bacterial genome sequence is likely to reduce in cost and increase in speed, it is foreseeable that a large number of complete bacterial genomes will be sequenced in the near future. A consequence of high throughput sequencing is a massive amount of genome sequence data, which in effect, is the ultimate data for Minimum SNPs analysis. Using Minimum SNPs, the total DNA sequence analysed for *S. aureus* MLST was 3198bp (Chapter Three) and this produced highly informative genotyping SNPs. Nonetheless, in comparison with an extensive database of diverse whole genome sequences, a Minimum SNPs analysis would, undoubtedly, identify a better set of key genotyping SNPs. This analysis, however, would result in inclusion of highly polymorphic regions, such as exotoxin gene families that would bias selection of SNPs from the accessory genome component. This is not desirable when searching for core genome SNPs. One way to circumvent this outcome would be to analyse a set of virtual chromosomes that only feature essential core genes, which Baba and co-workers have previously demonstrated is possible via comparison of the *S. aureus* Mu50, N315 and MW2 core genomes (Baba *et al.*, 2002). In addition to DNA sequence, complex data sets of binary marker variation can also be analysed using Minimum SNPs.

DNA microarrays are also whole genome characterisation methods; however instead of operating at the sequence level, microarrays determine the presence or absence of each gene across the genome. The results are essentially complete binary marker profiles that can be analysed using Minimum SNPs in a manner analogous to SCC*mec* binary markers. The analysis could be conducted in two ways: the first would be to identify a set of binary markers that differentiate the input profiles most efficiently, or second, by adding the complete Minimum SNPs binary profiles (As and Ts) to the core genome virtual chromosomes to select the best overall combination of targets. This approach is analogous to adding actual *mec*A SNPs to the SCC*mec* binary marker data set. Overall, by utilising the methods described here, the identification of efficient genotyping targets from an overwhelming mass of whole genome sequences and large databases of microarray binary marker data can be potentially, a simple and effective task.

Another major outcome of this study that can also be applied to *S. aureus* genotyping applications was the successful use of HRM to characterise the *spa* VNTR. HRM analysis proved to be an effective method for translating DNA sequence variation into genotype resolution and it is feasible that HRM has further applications for analysing additional VNTRs. A MLVA genotyping method featuring five VNTR loci has been developed for *S. aureus* and produces resolution equivalent with PFGE via size differentiation of multiplexed PCR amplicons (Sabat *et al.*, 2003). As each amplicon includes VNTR regions, it is feasible to convert the reactions to real-time PCR and then conduct HRM to produce genotyping results (Chapter 5). There are two ways in which this could be conducted. The first includes the use of unlabelled primers and SYBR green to amplify and perform HRM on the multiplex reaction simultaneously. The resulting melting profile would be a combination of the variation within each of the amplicons. If the amplicons encode different T_m values, then the normalised melt profile would display a staggered decrease in fluorescence as indicated by two or even three humps. These results could be analysed according to the earlier description (Chapter 5) and would potentially provide more discrimination power than the *spa* VNTR alone.

The second method would involve using fluorescently labelled primers for each reaction within the multiplex to produce individual HRM profiles. If the amplicon number exceeds the channels available, however, then multiple tubes would be required. Due to a requirement for labelled primers and multiple HRM curves, this method would be more expensive and require longer data interpretation time per isolate. Another point worth considering regarding HRM analysis is that longer amplicons are less informative compared with shorter amplicons. Furthermore, premix PCR reagents such as SYBR green are optimised for shorter fragments; therefore longer amplification reactions will require additional DNA polymerase, dNTPs and in some cases an increased concentration of $MgCl_2$. Although the conversion of MLVA to real-time PCR and HRM would possibly produce better resolution of genotypes, the issue remains that it is not anchored to a clonal method. Therefore, like the *spa* VNTR, it is recommended that a clonal genotyping method such as the high D SNPs be included with or without the SCC*mec* binary targets depending on the antibiotic resistance phenotype.

Real-time PCR and HRM analysis could also be utilised to determine the allelic identity of the high *D* SNP set. SNP genotyping using real-time PCR and HRM has been implemented in human genetics where not only the SNP state, but also the allelic ratio in pooled samples can be determined (Krypuy *et al.*, 2006). In the case of the high *D* SNP set, only small amplicons crossing each of the seven SNPs would be required, which could then be analysed by HRM to determine the allelic state in the same manner that the *spa* VNTR was examined. HRM analysis could potentially produce differential results for each allelic state more time efficiently and cost effectively than allele-specific real-time PCR. HRM analysis of the high *D* SNP set could easily be added to the established *spa* VNTR assay and potentially to other VNTR loci. Furthermore, the SCC*mec* binary targets could be included to produce a simplified version of the PHRANA strategy detailed here.

One major drawback of SNP analysis using HRM is that non-target SNPs lying within the amplicon will affect the outcome. While this can be accounted for in the results interpretation, even if the SNP is novel, fragments that intentionally feature more than one SNP may actually be advantageous if the allelic combination can be accounted for. Therefore, it is possible to characterise the seven high *D* SNPs in five amplicons completely since SNP pairs *tpi*241, *tpi*243 and *arc*C162, *arc*C210 are co-amplifiable. It is also worth mentioning that these reactions can be multiplexed using fluorescent primers, therefore *S. aureus* high *D* SNP genotyping could be reduced to two real-time PCR tubes. Given a reduction in number of reactions and tubes, the addition of important SNPs, such as for ST93, is a practical exercise. Overall, real-time PCR and HRM analysis is an important emerging tool for bacterial genotyping applications, as it provides a cost effective and rapid method capable of accurately differentiating DNA polymorphisms.

The methods developed here were based around the real-time PCR platform. This is because they are generic, moderately-priced instruments that are capable of interrogating different classes of polymorphisms, including SNPs, binary markers and now VNTR regions. In addition, real-time PCR methods are single step and closed tube, making them very attractive since they reduce contamination in the clinical laboratory, and do not depend on batching of assays for cost effectiveness as are

competing technologies such as reverse line blots (Kong *et al.*, 2005). However, there are other experimental platforms that could provide practical alternatives.

High throughput DNA sequence analysis methods have undergone considerable advances in recent times. A prime example is the 454 micro-bead sequencing methodology that produces very large amounts of sequence data via parallel pyrosequencing reactions. For whole genome sequencing, the method is broadly based on creation of a single stranded library of random 300-800bp fragments that are ligated to DNA adapters featuring amplification and sequencing primer sites. For sequencing individual specified genomic regions, small PCR fragments can also be used. Capture beads coated in the complementary primer are then used to capture a single fragment that is then added to a water and oil emulsion to produce exclusive PCR chambers for each fragment. After the amplification process that results in the capture beads being covered with hundreds of thousands of fragment copies, the beads are enriched and treated to produce single stranded fragments that are ready for sequencing. A picotitre plate featuring more than 400K wells is used to provide a single well for each bead during the simultaneous pyrosequencing reaction. Pyrosequencing involves the sequential analysis of each nucleotide after it is washed across the picotitre plate. Each incorporated nucleotide results in a fluorescent reaction that is recorded individually by fibre optics for each well. This process continues until the reaction is exhausted. Computer software is then utilised to compose the individual fragments into a genome sequence that unlike the standard Sanger method, requires a references genome sequence to complete. The first complete *S. aureus* genome sequenced by this method was published recently (Highlander *et al.*, 2007).

The main technical drawback of 454 sequencing applications to microbial genotyping is that the current double stranded sequencing reads are limited to about 200bp. If single strand sequencing is adequate, then a read of about 400bp can be achieved, however, this is still not long enough for *S. aureus* MLST. This restriction could be easily overcome by amplification of the MLST loci followed by random fragmentation as is conducted when preparing genomes for sequencing. MLST using this method could provide a rapid and high throughput alternative to the current sequencing protocol. Although resolution would be increased by analysis of the entire

163

MLST loci, sequence analysis across the site of the high D SNPs is also a possibility. In comparison with full 454 MLST, this has the benefit of only sequencing five small fragments, as two pairs of high D SNPs are located close to one another. It is predicted, however, that the maximal 454 sequencing length will be extended to about 500bp in the near future, which would increase the suitability of this method for full MLST analysis.

The use of 454 sequencing could also be used to simultaneously analyse other targets identified in the three tiered PHRANA strategy. Detection of SCC*mec* and virulence binary markers is potentially a simple matter due to presence or absence of the targets. Analysis of the polymorphic *ccr* sequence could also be included to increase genotyping resolution as has been demonstrated previously (Oliveira *et al.*, 2006b). Furthermore, use of high throughput sequencing also allows inclusion of sequence analysis of the right SCC*mec* extremity that has also been shown to exhibit useful variation (Cuny and Witte, 2005). Analysis of the *spa* VNTR, or other VNTRs, is also potentially applicable to the 454 sequencing platform. The success of this application will, however, depend on the length of sequencing read, as 454 sequencing has noted difficulties when dealing with repeat region sequence. Overall, 454 sequencing and other high throughput sequencing platforms such as Solenix appear potentially useful for analysing a range of polymorphisms that are currently exploited in bacterial genotyping.

To be considered an economical alternative to the current sequencing methods for bacterial genotyping, the 454 method must be able to analyse and differentiate pooled samples featuring large numbers of isolates simultaneously. Until recently, the analysis of pooled samples was restricted to human genetic studies based on determining the allelic frequency of specific traits whereby individual samples remain anonymous. A new method termed parallel tagged sequencing (PTS) is, however, based around ligating a sample specific short oligo to the fragment before the ligation of the 454 primer adapter (Meyer *et al.*, 2007). This allows sample identification during the post sequence analysis, an approach although not yet applied, that would be highly amenable to 454 bacterial genotyping. As more than 400k wells are available per run; the number of samples is limited to the number of unique identifiers

that could potentially allow 454 MLST or high D SNP sequencing to become an affordable method in the future.

The largest barrier for 454 and other high throughput DNA sequencing methods that prohibits its use for routine bacterial genotyping is the large capital investment and cost per sequencing run. Although practical for whole genome sequencing, the analysis of 400k fragments is well above the requirement for bacterial genotyping, which does not require an abundance of comparative data to produce high resolution genotypes. It is tantalising to consider, however, the future possibility of a specialised picotitre plate and 454 machine purposely designed for bacterial genotype marker analysis. If such a machine could be produced and operated economically, then it would provide an ideal platform for high throughput bacterial genotyping.

Another high throughput platform that could potentially be used for bacterial genotyping is matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry (MALDI-TOF MS). The DNA sequencing capabilities of MALDI-TOF MS is currently 100bp per read and can be achieved at high throughput with low reaction cost. There are a number of preparatory methods for MALDI-TOF MS DNA sequencing; however, the end result is a sample containing a ladder of DNA sequences with each fragment displaying different masses relative to its length. The MALDI-TOF MS analysis differentiates the fragments based on the mass differences and determines the nucleotide identity via specific fluorescent detection that is analogous to DNA sequencing using capillary electrophoresis. Due to short reads, this DNA sequencing method is suitable for bacterial genotyping via SNP detection, rather than full MLST. The detection of binary markers is also a possibility, but it is not suitable for VNTR sequencing until the read length is increased considerably. Although advances are being made to increase length, it is not certain what role MALDI-TOF MS will play in the DNA sequencing market in the future.

MALDI-TOF MS can also act as a resequencing method for short DNA fragments that differ in sequence mass resulting from polymorphisms (Bray *et al.*, 2001). In effect, this process can be used for bacterial genotyping in an analogous manner to HRM (Chapter Five). This MALDI-TOF MS application is well established as demonstrated by commercialisation of integrated MS machines with custom

developed software for DNA SNP analysis. One example is the iPLEX system developed by Sequenom, which can be used to analyse 384 samples per plate with up to 36 pooled fragments per sample. Multiplexing capability is advantageous for analysis of multiple loci, such as the high *D* SNPs set. This analysis could be achieved by multiplex amplification of the seven MLST loci, followed by single nucleotide extension of a primer flanking the high *D* SNP using a termination nucleotide mix. The mass of the resulting small fragments will be determined by the termination nucleotide incorporated and will ultimately allow MALDI-TOF MS to differentiate alleles. This SNP analysis is also applicable for key *ccr* and *mec*A SNPs that could be used to extract genotyping information from the SCC*mec* element. In comparison with the real-time PCR platform, MALDI-TOF MS achieves superior throughput of multiplexed SNPs, making it well suited to large scale investigations. MALDI-TOF MS requires, however, multiple experimental steps per isolate and is hindered by a largely prohibitive start-up cost.

In contrast to the complex and expensive DNA analysis platforms, there is a desire for simple and cost effective point-of-care diagnostics platforms that are commonly referred to as "lab-on-a-chip" (LOaC) devices. LOaC devices aim to miniaturise common laboratory processes on small chips that incorporate the thermal, electrical, and fluidic requirements for the experiment. LOaC devices have been designed to fulfil a wide range of applications, one being the real-time PCR detection of DNA sequence including HRM (Neuzil *et al.*, 2006). Although still in their infancy, it is possible that future LOaC designs could feature numerous reaction tubes or multiplexing capabilities that enable simultaneous analysis of a number of polymorphisms. A MRSA genotyping point-of-care device using HRM to genotype the five high *D* fragments and the *spa* VNTR locus, in addition to a small selection of SCC*mec* binary targets would be very useful for hospital applications. As hypersensitive LOaC devices can function on a miniscule amount of sample template, there is also the possibility of bacterial detection and genotyping directly from clinical samples, in a single step. The LOaC devices have the potential to be effective future methods for detection and genotyping of infectious bacteria.

The recent explosion of comparative genomic data production and bioinformatic research has combined with advances in laboratory equipment technology to usher in

an era of computer-based, advanced bacterial genomic analysis. Comparative methods including high-throughput DNA sequencing and microarray analysis have resulted in an unprecedented and rapidly expanding quantity of data. Therefore it is imperative that computer assisted methods are developed to exploit the information in this data maximally, for the purpose of developing rapid diagnostic methods for infectious disease causing organisms.

The current study has described a comprehensive three tiered MRSA genotyping strategy whereby genotypes are generated by characterisation of multiple loci that exhibit progressively higher mutation rates. Genotyping targets for the first two tiers were optimised using a computerised approach and the third tier was interrogated using a novel application of HRM. With utilisation of the strategy described here, it is possible to genotype an unknown MRSA isolate and position it within the global MLST population structure, determine the SCC*mec* type, and generate a *spa* genotype that together produce a robust combined genotype. Moreover, it is possible to generate a comprehensive genotype on standard laboratory equipment using standard reagents in a cost and time effective manner, traits that make this novel MRSA genotyping strategy ideally suited for routine clinical use.

APPENDIX

Appendix Table 1. 107 MRSA isolates used in this study.

AGAR 2000 community onset collection*	ST	SCC*mec*	*agr*	comments	PFGE	PVL	SdrE or Bbp	SdrE only	Cna	tst	fnbA	IcaA	hlg
PC 8	1	4a	3		ND	0	1	1	1	0	1	1	1
SJOG 30	1	4a	3		ND	1	1	1	1	0	1	1	1
IMVS 67	8	5	1		ND	0	1	1	0	0	1	1	1
RPH 2	8	new	1	mec class B no ccr	ND	0	1	1	1	0	1	1	1
CH 16	22	4	1		ND	0	1	1	1	0	1	1	1
PAH 1	30	4	3		ND	1	1	0	1	0	1	1	1
PAH 58	30	4a	3		ND	1	1	0	1	0	1	1	1
RPH 74	45	5	1		ND	0	1	1	1	1	1	1	1
CH 97	73	4a	2		ND	0	1	1	1	0	1	1	1
RBH 98	93	4a	3		ND	1	1	1	0	0	1	1	1
AH 1	ST128 (snp239)	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
AH 13	239	3	1	pT181 + pI258	ND	0	0	0	1	0	1	1	1
AH 41	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
CH 10	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
FMC 1	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
FMC 19	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
GP 35 (sa)	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
IMVS 20	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
IMVS 51	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
RCH 66	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
RDH 81	239	3	1	pT181 + pI258	ND	0	1	1	1	0	1	1	1
LVH 25	239	3 (new)	1	pUB110	ND	0	1	1	1	0	1	1	1
RPAH 15	239	3 (new)	1	pUB110	ND	0	1	1	1	0	1	1	1
RPAH 18	239	3 (new)	1	pUB110	ND	0	1	1	1	0	1	1	1
FH 43	~1	4a	3		ND	0	1	1	1	0	1	1	1
FH 53	~1	1	3	mec class B + ccr A1	ND	0	1	1	1	0	1	1	1
FMC 53	~1	4a	3		ND	0	1	1	1	0	1	1	1
PC 13	~1	4a	3		ND	0	1	1	1	0	1	1	1
RDH 69	~1	4a	3		ND	0	1	1	1	0	1	1	1
RHH 10	~1	4a	3		ND	0	0	0	1	0	1	1	1
RHH 58	~1	4a	3		ND	0	0	0	1	0	1	1	1
RNSH 38	~1	4a	3		ND	0	1	1	1	0	1	1	1
RPH 29	~1	4a	3		ND	0	1	1	1	0	1	1	1
RPH 3	~1	4a	3		ND	0	1	1	1	0	1	1	1
RPH 32	~1	4a	3		ND	0	1	1	1	0	1	1	1
RPH 85	~1	4a	3		ND	1	1	1	1	0	1	1	1
SJOG 29	~1	4a	3		ND	0	1	1	1	0	1	1	1
SN 39	~1	new	3	mec class C + ccr A1+ pT181	ND	0	1	1	1	0	1	1	1
SN 49	~1	new	3	mec class C + ccr A1+ pT181	ND	0	1	1	1	0	1	1	1
CH 45	~22	4	1		ND	0	1	1	1	0	1	1	1
CH 69	~22	4	1		ND	0	1	1	1	0	1	1	1
CH 74	~22	4	1		ND	0	1	1	1	0	1	1	1
CH 39	~30	4a	3		ND	1	1	0	1	0	1	1	1
CH 94	~30	4a	3		ND	1	1	0	1	0	1	1	1
GPI	~30	4a	3		ND	1	1	0	1	0	1	1	1
NH 100	~30	4a	3		ND	1	1	0	1	0	1	1	1
WEST 40	~30	4a	3		ND	1	1	0	1	0	1	1	1

Appendix Table 1 continued.

^Nimmo *et al.* 2000 South east QLD	ST	SCC*mec*	*agr*	comments	PFGE	PVL	SdrE or Bbp	SdrE only	Cna	tst	fnbA	IcaA	hlg
E804531	5 (snp73)	4a	2		I	0	1	1	0	0	1	1	1
A803355	30	4a	3		A0	1	1	0	1	0	1	1	1
66460/98	30	4a	3		A0	1	1	0	1	0	1	1	1
MA806533	30	4a	3		A0	1	1	0	1	0	1	1	1
IP01M1081	ST59 (snp93)	4a	1		Q	0	1	1	0	0	1	1	1
IP01M2046	78	4a	3		P1	0	1	1	1	0	1	1	1
MC801535	88	new	3	mec class B no ccr	D	0	1	1	0	0	1	1	1
F829549	88	4	3		D	0	1	1	0	0	1	1	1
MD828534	~1	new	3	mec class C + ccr A1+ pT181	E	0	1	1	1	0	1	1	1
MB827549	~1	new	3	mec class C + ccr A1+ pT181	E	0	1	1	1	0	1	1	1
MD808118	~1	new	3	mec class C + ccr A1+ pT181	L	0	1	1	1	0	1	1	1
B8-10	~8	4a	1		ND	0	1	1	0	0	1	1	1
D828570	~30	4a	3		A0	1	1	0	1	0	1	1	1
E822547	~30	4	3		A0	1	1	0	1	0	1	1	1
E802537	~30	4a	3		A3	1	1	0	1	0	1	1	1
MA823547	~30	4a	3		A1	1	1	0	1	0	1	1	1
F810539	~30	4a	3		A0	1	1	0	1	0	1	1	1
MD821552	~30	4a	3		A2	1	1	0	1	0	1	1	1
MC810534	~30	4a	3		A1	1	1	0	1	0	1	1	1
MB826559	~30	4	3		A0	1	1	0	1	0	1	1	1
E803534	~30	4a	3		A0	1	1	0	1	0	1	1	1
MD817541	~30	4a	3		A0	1	1	0	1	0	1	1	1
MA830538	~30	4a	3		A0	1	1	0	1	0	1	1	1
I825560	~30	4	3		A0	1	1	0	1	0	1	1	1
68284/98	~30	4	3		A5	1	1	0	1	0	1	1	1
E822485	ST36 (snp30)	2	3	pUB110	B	0	1	0	1	1	1	1	1
MJ710566	~45	5	1		C	0	1	1	1	1	1	1	1
B8-31	~45	5	1		K	0	1	1	1	1	1	1	1
K704540	~239	3a	1	pI258	F	0	1	1	1	0	1	1	1
K705613	~239	3a	1	pI258	F2	0	1	1	1	0	1	1	1
K711532	~239	3a	1	pI258	F3	0	1	1	1	0	1	1	1
K714372	~239	3a	1	pI258	F4	0	1	1	1	0	1	1	1
MIC703484	~239	3 (new)	1	pUB110	G1	0	1	1	1	0	1	1	1
I823541	~239	3 (new)	1	pUB110	G2	0	1	1	1	0	1	1	1
E812560	~239	3 (new)	1	pUB110	J	0	1	1	1	0	1	1	1
IPOOM14235	~239	3 (new)	1	pUB110	O	0	1	1	1	0	1	1	1

Appendix Table 1 continued.

Nimmo, Darling Downs community outbreak, 2004	ST	SCC*mec*	agr	comments	PFGE	PVL	SdrE or Bbp	SdrE only	Cna	tst	fnbA	IcaA	hlg
15952-7223	~1	4a	3		ND	1	1	1	1	0	1	1	1
15069-6625	~30	4	3		ND	1	1	0	1	0	1	1	1
16516-6076	~30	4a	3		ND	1	1	0	1	0	1	1	1
16077-5584	~93	4a	3		ND	1	1	1	0	0	1	1	1
15817-3897	~93	4a	3		ND	1	1	1	0	0	1	1	1
16530-0503	~93	4a	3		ND	1	1	1	0	0	1	1	1
16530-3565	~93	4a	3		ND	1	1	1	0	0	1	1	1
15808-0020	~93	4a	3		ND	1	1	1	0	0	1	1	1
15816-9412	~93	4a	3		ND	1	1	1	0	0	1	1	1
15821-5377	~93	4a	3		ND	1	1	1	0	0	1	1	1
15821-6052	~93	4a	3		ND	1	1	1	0	0	1	1	1
16076-8046	~93	4a	3		ND	1	1	1	0	0	1	1	1
16077-1053	~93	4a	3		ND	1	1	1	0	0	1	1	1
16279-9349	~93	4a	3		ND	1	1	1	0	0	1	1	1
15814-9852	~239	3 (new)	1	pUB110	ND	0	1	1	1	0	1	1	1

Munckhof, South East QLD, 2004	ST	SCC*mec*	agr	comments	PFGE	PVL	SdrE or Bbp	SdrE only	Cna	tst	fnbA	IcaA	hlg
IPOOM11247	~30	4	3		ND	1	1	0	1	0	1	1	1
IPOOM3393	~30	4	3		ND	1	1	0	1	0	1	1	1
IP2001M1984	~30	4	3		ND	1	1	0	1	0	1	1	1
14982-3245	~30	4	3		ND	1	1	0	1	0	1	1	1
13792-4492	~93	4a	3		ND	1	1	1	0	0	1	1	1
13840-1510	~93	4a	3		ND	1	1	1	0	0	1	1	1
14588-6938	~93	4a	3		ND	1	1	1	0	0	1	1	1
14622-7928	~93	4a	3		ND	1	1	1	0	0	1	1	1
14176-5710	~239	3b	1	no plasmids	ND	0	1	1	1	0	1	1	1

1 = positive 0 = negative

Isolates in the "ST" column not labelled with "~" were subject to full MLST and correspond to the known isolates from 3.3.2.

Appendix Table 2. 25 GenBank *mec*A sequence files

SCC*mec*	GenBank #	Strain	*mec*A SNP profile*
1B.1.1	AB033763	NCTC10442	ACGTGTT**A**
1B.1.1	CP000046	COL	ACGTGTT**A**
2A.1.1	BA000018	N315	**C**CGTGTTG
2A.1.1	BX571856	MRSA252	ACGTGTTG
2A.1.1	NC_002758	Mu50	**C**CGTGTTG
2A.3.5	AJ810120	AR13.1_3330.2	ACGTGTTG
2B.1	AB063172	CA05	ACGTGTTG
2B.1	AB221121	80s-3	ACGTGTTG
2B.1	AB221122	80s-4	ACGTGTTG
2B.1	CP000255	USA300	ACGTGTTG
2B.1	NC_003923	MW2	ACGTGTTG
2B.2.1	AB063173	JCSC1978	ACGTGTTG
2B.3.1	AB096217	MR108	ACGTGTT**A**
2B.3.1	AB221120	80s-2	ACGTGTT**A**
2B.3.1	AB221124	NN1	ACGTGTTG
2B.3.1	AB236888	90s-2	ACGTGTT**A**
2B.3.1	AB245470	Na	ACGTGTTG
2B.3.1	AB245471	Na	ACGTGTT**A**
2B.3.2	AY271717	2314	ACGTGTT**A**
2B.3.3	AJ810121	AR43_3330.1	ACGTGTTG
2B.5	DQ106887	M03-68	ACGTGTT**A**
3A.1.1	AB037671	85_2082	ACGTG**GT**A
3A.1	AB047089	85_3907	ACGTGTT**A**
5C.1	AB121219	WIS	ACGTGT**A**G
6C2.VT	AY894415	TSGH17	ACGTGT**A**G

**mec*A SNP positions from 3` end - 75, 312, 415, 438, 448, 612, 675, 737.

Appendix Table 3. 19 *mec*A from this study

SCC*mec*	Strain	*mec*A SNP profile*	SNP/MLST clonal complex
1B.1.1	FH53	ACGTGTTG	1
1C	MB827549	ACGTGTAG	1
1C	SN49	ACGTGTAG	1
2A.1.1	E822485	ACGTGTTG	36
2B.1	FH43	ACGTGTTG	1
2B.1	24	ACGTGTTG	5
2B.1	E804531	ACGTGTTG	5
2B.1	IP01M1081	ACGTGTTG	59
2B.1	NH100	ACGTGTTG	30
2B.1	RBH98	ACGTGTTG	93
2B.1	SJOG30	ACGTGTTG	30
3A.1.(5)	15814-9852	ACATGTTG	239
3A.1.(5)	LVH25	ACGTGTTG	239
3A.1.1	AH13	ACGAGGTA	239
3A.1.1	CH10	ACGTAGTA	239
3A.1.2	K704540	AAGTGGTA	239
3A	14176-5710	ACGTGTTG	239
5C.1	IMVS67	ACGTGTAG	8
5C.1	MJ710566	ACGTGTAG	45

**mec*A SNP positions from 3` end - 75, 312, 415, 438, 448, 612, 675, 737.

Identification of the pvl Bacteriophage by Southern Blot Analysis

The *pvl* toxin is associated with CA-MRSA and the severe condition of necrotising pneumonia. Each of the CA-MRSA isolates from the SNP ST30 and 93 profiles were found to be positive for *pvl*, however, isolates from the SNP ST1 profile or otherwise known as the WAMRSA-1 clone were found to contrast in their *pvl* carriage. The uneven distribution of the *pvl* carriage is facilitated by bacteriophage transduction. A number of phage are known to carry *pvl*, however, the exact vehicle for transmission in Australian has yet to be elucidated. The primary purpose of identifying the *pvl* phage via its insertion location was to determine whether ST1 *pvl* positive and negative isolates could be differentiated using PFGE restriction profiles. Furthermore *pvl* phage identification may also serve to increase genotyping resolution if different phage types circulate within clones. The location of the *pvl* phage insertion site was determined by firstly genotyping five *pvl* positive and seven *pvl* negative isolates using PFGE, then detecting the phage location by southern blot analysis using a *pvl* PCR amplicon as probe.

The southern blot revealed that the *pvl* toxin phage was located within the largest PFGE fragment within each of the *pvl* positive isolates (Appendix Fig 1). As the PFGE fragment encoding *pvl* is over 1Mbps it was virtually impossible using standard PFGE to determine whether a ST1 isolate harbours the 40Kb *pvl* phage or not. A virtual restriction digest of the MW2 ST1 genome confirmed that this fragment harbours an attachment site and *pvl* phage SLT. As each ST1 isolate displayed the identical results it was concluded that the *pvl* gene was carried by the same phage. This was confirmed by bridging PCR from the genome into phage SLT.

A
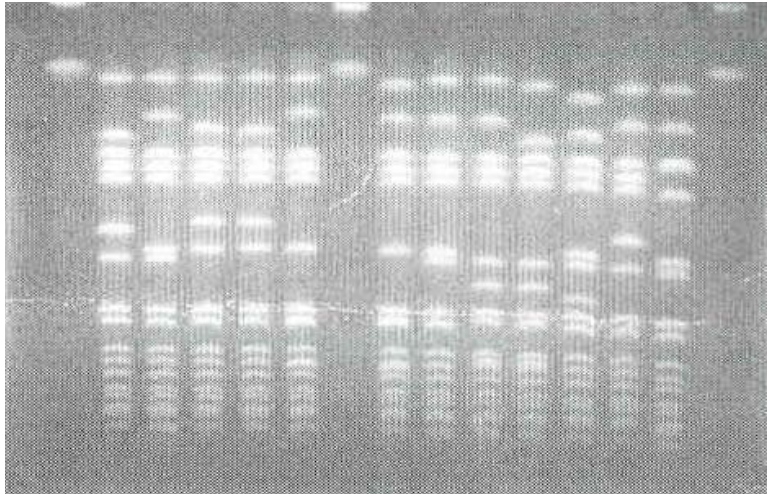
B
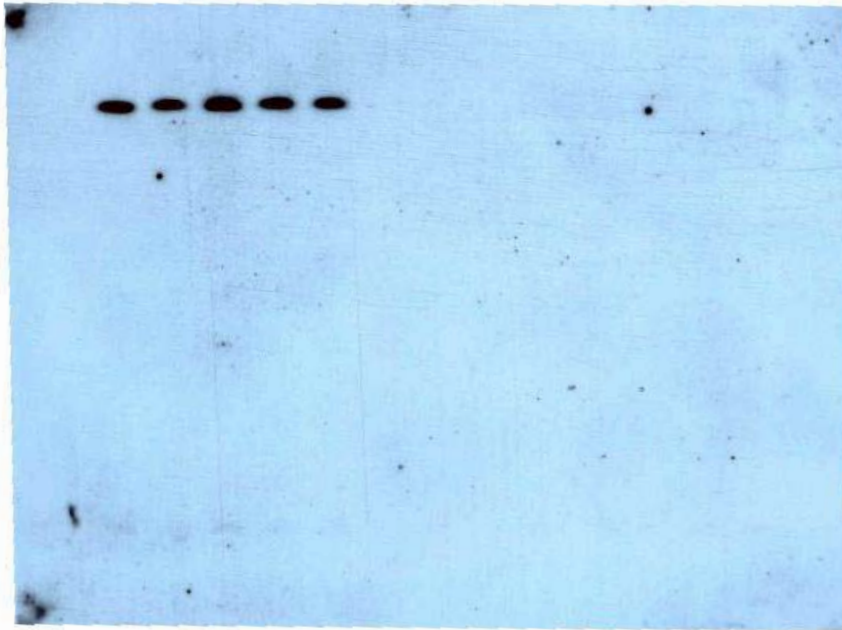
Appendix Fig 1. Southern blot result for ST1 *pvl* gene. A) The PFGE gel of the five *pvl* positive isolates (left lanes) and the seven ST1 *pvl* negative isolates (right lanes). B) Southern blot result when probing with a *pvl* amplicon shows that the largest fragment harbours the *pvl* phage. No obvious differences appear between the *pvl* positive and negative PFGE profiles.
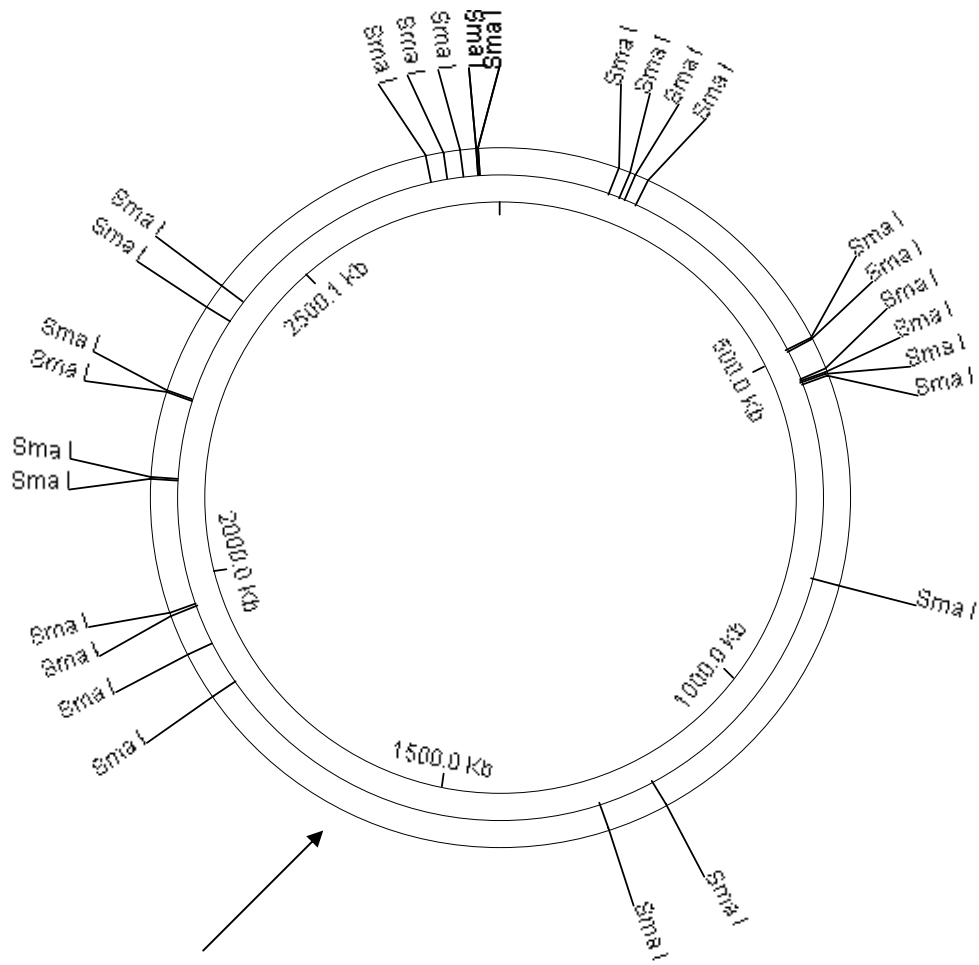
Appendix Fig 2. A virtual *sma*1 digest of the ST1 MW2 genome. The arrow indicates the fragment where the *pvl* phage had inserted.

Appendix Table 4. Unique sequences for SCC*mec* binary targets

SCC*mec* targets	NCBI reference	Characteristic Nucleotide Sequence
cassette chromosome recombinase A1 (*ccr*A1)	AB033763	23692..25041
cassette chromosome recombinase A2 (*ccr*A2)	BA000017	65008..66357
cassette chromosome recombinase 5 (*ccr*5)	AB121219	16132..17754
cassette chromosome recombinase C-VT (*ccr*C-VT)	AY894416	2921..4600
mec class A	BA000017	46999..49226
mec class A.3	AJ810120	16341..20315
mec class B	AB033763	28473..31389
mec class B1*	No sequence data	No sequence data
mec class C2	AB121219	8030..9439
Jya (Junkyard a or J1 region of 2B.1)	AB063172	4281..5771
Jyb (Junkyard b or J1 region of 2B.2.1)	AB063173	2195..2310
Jyc (Junkyard c or J1 region of 2B.3.1)	AB096217	8625..9280
IVd unique	AB097677	2622..3609
IVg unique	DQ106887	196..1944
pT181	AB037671	29641..30000
pUB110	AF181950	7021..8821
Tn*554* MLS	AJ810120	10361..11040
phi Tn*554* cad	AJ810120	9001..11100
Tn*4001*	AB096217	27001..27660
dcs	AB063172	23341..24300
Downstream pls locus A**	AB033763	18398..18871
3A.2.1 unique	AB047088	9241-9420 and 12600-12780***
IS256 in dcs	AB096217****	25801..26520

*(Lim *et al.*, 2003)

** Multiplex PCR for SCC*mec* locus A (Oliveira and de Lencastre, 2002).

*** A PCR between these regions is unique for 3A.2.1. Although unique regions for this element have been reported, no sequence data is available.

**** IS256 dcs insertion in IV.3 subtype.

*ccr*5 is utilised with *ccr*1 for the nucleated analysis

LITERATURE CITED

Aires De Sousa, M., Conceicao, T., Simas, C. & De Lencastre, H. (2005) Comparison of genetic backgrounds of methicillin-resistant and -susceptible *Staphylococcus aureus* isolates from Portuguese hospitals and the community. *J Clin Microbiol,* 43**,** 5150-7.

Aires De Sousa, M. & De Lencastre, H. (2003) Evolution of sporadic isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) in hospitals and their similarities to isolates of community-acquired MRSA. *J Clin Microbiol,* 41**,** 3806-15.

Allegra, G., Niutta, R. & Giuffrida, G. (1954) [Studies on synthesis of carotene and on the role of carotenoids in *Staphylococcus aureus*. I. Effect of glucose on chromogenesis of *Staphylococcus aureus* (Oxford strain).]. *Riv Ist Sieroter Ital,* 29**,** 263-9.

Anonymous (1999) From the Centers for Disease Control and Prevention. Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus*--Minnesota and North Dakota, 1997-1999. *Jama,* 282**,** 1123-5.

Anonymous (2002) *Staphylococcus aureus* resistant to vancomycin--United States, 2002. *MMWR Morb Mortal Wkly Rep,* 51**,** 565-7.

Archer, G. L., Niemeyer, D. M., Thanassi, J. A. & Pucci, M. J. (1994) Dissemination among staphylococci of DNA sequences associated with methicillin resistance. *Antimicrob Agents Chemother,* 38**,** 447-54.

Armstrong-Esther, C. A. (1976) Carriage patterns of *Staphylococcus aureus* in a healthy non-hospital population of adults and children. *Ann Hum Biol,* 3**,** 221-7.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. (2008) Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol,* 190**,** 300-10.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. & Hiramatsu, K. (2002) Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet,* 359**,** 1819-27.

Baillargeon, J., Kelley, M. F., Leach, C. T., Baillargeon, G. & Pollock, B. H. (2004) Methicillin-resistant *Staphylococcus aureus* infection in the Texas prison system. *Clin Infect Dis,* 38**,** e92-5.

Bannerman, T. L., Hancock, G. A., Tenover, F. C. & Miller, J. M. (1995) Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of *Staphylococcus aureus*. *J Clin Microbiol,* 33**,** 551-5.

Bartels, M., Boye, K., Rhod Larsen, A., Skov, R. & Westh, H. (2007) Rapid Increase of Genetically Diverse Methicillin-Resistant *Staphylococcus aureus*, Copenhagen, Denmark. *Emerg Infect Dis,* 13**,** 1533-1540.

Bayles, K. W. & Iandolo, J. J. (1989) Genetic and molecular analyses of the gene encoding staphylococcal enterotoxin D. *J Bacteriol,* 171**,** 4799-806.

Beck, W. D., Berger-Bachi, B. & Kayser, F. H. (1986) Additional DNA in methicillin-resistant *Staphylococcus aureus* and molecular cloning of *mec*-specific DNA. *J Bacteriol,* 165**,** 373-8.

Begier, E. M., Frenette, K., Barrett, N. L., Mshar, P., Petit, S., Boxrud, D. J., Watkins-Colwell, K., Wheeler, S., Cebelinski, E. A., Glennen, A., Nguyen, D. & Hadler, J. L. (2004) A high-morbidity outbreak of methicillin-resistant

Staphylococcus aureus among players on a college football team, facilitated by cosmetic body shaving and turf burns. *Clin Infect Dis,* 39**,** 1446-53.

Bergdoll, M. S., Crass, B. A., Reiser, R. F., Robbins, R. N. & Davis, J. P. (1981) A new staphylococcal enterotoxin, enterotoxin F, associated with toxic-shock-syndrome *Staphylococcus aureus* isolates. *Lancet,* 1**,** 1017-21.

Bibel, D. J., Aly, R., Shinefield, H. R., Maibach, H. I. & Strauss, W. G. (1982) Importance of the keratinized epithelial cell in bacterial adherence. *J Invest Dermatol,* 79**,** 250-3.

Blair, J. E. & Carr, M. (1960) The techniques and interpretation of phage typing of staphylococci. *J Lab Clin Med,* 55**,** 650-62.

Blanc, D. S., Francioli, P. & Hauser, P. M. (2002) Poor value of pulsed-field gel electrophoresis to investigate long-term scale epidemiology of methicillin-resistant *Staphylococcus aureus. Infect Genet Evol,* 2**,** 145-8.

Blanc, D. S., Petignat, C., Moreillon, P., Entenza, J. M., Eisenring, M., Kleiber, H., Wenger, A., Troillet, N., Blanc, C. & Francioli, P. (1999) Unusual spread of a penicillin-susceptible methicillin-resistant *Staphylococcus aureus* clone in a geographic area of low incidence. *Clin Infect Dis,* 29**,** 1512-8.

Boost, M. V., O'donoghue, M. M. & Siu, K. H. (2007) Characterisation of methicillin-resistant *Staphylococcus aureus* isolates from dogs and their owners. *Clin Microbiol Infect,* 13**,** 731-3.

Boyce, J. M. (1989) Methicillin-resistant *Staphylococcus aureus*. Detection, epidemiology, and control measures. *Infect Dis Clin North Am,* 3**,** 901-13.

Boye, K., Bartels, M. D., Andersen, I. S., Moller, J. A. & Westh, H. (2007) A new multiplex PCR for easy screening of methicillin-resistant *Staphylococcus aureus* SCC*mec* types I-V. *Clin Microbiol Infect,* 13**,** 725-7.

Boyle-Vavra, S., Ereshefsky, B., Wang, C. C. & Daum, R. S. (2005) Successful multiresistant community-associated methicillin-resistant *Staphylococcus aureus* lineage from Taipei, Taiwan, that carries either the novel Staphylococcal chromosome cassette *mec* (SCC*mec*) type VT or SCC*mec* type IV. *J Clin Microbiol,* 43**,** 4719-30.

Bray, M. S., Boerwinkle, E. & Doris, P. A. (2001) High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat,* 17**,** 296-304.

Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A,* 83**,** 3746-50.

Brigido Mde, M., Barardi, C. R., Bonjardin, C. A., Santos, C. L., Junqueira, M. L. & Brentani, R. R. (1991) Nucleotide sequence of a variant protein A of *Staphylococcus aureus* suggests molecular heterogeneity among strains. *J Basic Microbiol,* 31**,** 337-45.

Brussow, H., Canchaya, C. & Hardt, W. D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev,* 68**,** 560-602.

Centofanti, P., Savia, F., La Torre, M., Ceresa, F., Sansone, F., Veglio, V., Fossati, L., Guglielmi, E. & Rinaldi, M. (2007) A prospective study of prevalence of 60-days postoperative wound infections after cardiac surgery. An updated risk factor analysis. *J Cardiovasc Surg (Torino),* 48**,** 641-6.

Cha, H. Y., Moon, D. C., Choi, C. H., Oh, J. Y., Jeong, Y. S., Lee, Y. C., Seol, S. Y., Cho, D. T., Chang, H. H., Kim, S. W. & Lee, J. C. (2005) Prevalence of the ST239 clone of methicillin-resistant *Staphylococcus aureus* and differences

in antimicrobial susceptibilities of ST239 and ST5 clones identified in a Korean hospital. *J Clin Microbiol,* 43**,** 3610-4.

Chambers, H. F. (2001) The changing epidemiology of *Staphylococcus aureus*? *Emerg Infect Dis,* 7**,** 178-82.

Charlebois, E. D., Bangsberg, D. R., Moss, N. J., Moore, M. R., Moss, A. R., Chambers, H. F. & Perdreau-Remington, F. (2002) Population-based community prevalence of methicillin-resistant *Staphylococcus aureus* in the urban poor of San Francisco. *Clin Infect Dis,* 34**,** 425-33.

Choi, Y., Lafferty, J. A., Clements, J. R., Todd, J. K., Gelfand, E. W., Kappler, J., Marrack, P. & Kotzin, B. L. (1990) Selective expansion of T cells expressing V beta 2 in toxic shock syndrome. *J Exp Med,* 172**,** 981-4.

Chongtrakool, P., Ito, T., Ma, X. X., Kondo, Y., Trakulsomboon, S., Tiensasitorn, C., Jamklang, M., Chavalit, T., Song, J. H. & Hiramatsu, K. (2006) Staphylococcal cassette chromosome *mec* (SCC*mec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for SCC*mec* elements. *Antimicrob Agents Chemother,* 50**,** 1001-12.

Chu, G., Vollrath, D. & Davis, R. W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science,* 234**,** 1582-5.

Chung, M., De Lencastre, H., Matthews, P., Tomasz, A., Adamsson, I., Aries De Sousa, M., Camou, T., Cocuzza, C., Corso, A., Couto, I., Dominguez, A., Gniadkowski, M., Goering, R., Gomes, A., Kikuchi, K., Marchese, A., Mato, R., Melter, O., Oliveira, D., Palacio, R., Sa-Leao, R., Santos Sanches, I., Song, J. H., Tassios, P. T. & Villari, P. (2000) Molecular typing of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis: comparison of results obtained in a multilaboratory effort using identical protocols and MRSA strains. *Microb Drug Resist,* 6**,** 189-98.

Clewell, D. B., An, F. Y., White, B. A. & Gawron-Burke, C. (1985) *Streptococcus faecalis* sex pheromone (cAM373) also produced by *Staphylococcus aureus* and identification of a conjugative transposon (Tn918). *J Bacteriol,* 162**,** 1212-20.

Cohen, P. R. (2007) Community-acquired methicillin-resistant *Staphylococcus aureus* skin infections: a review of epidemiology, clinical features, management, and prevention. *Int J Dermatol,* 46**,** 1-11.

Cole, A. M., Dewan, P. & Ganz, T. (1999) Innate antimicrobial activity of nasal secretions. *Infect Immun,* 67**,** 3267-75.

Cole, A. M., Tahk, S., Oren, A., Yoshioka, D., Kim, Y. H., Park, A. & Ganz, T. (2001) Determinants of *Staphylococcus aureus* nasal carriage. *Clin Diagn Lab Immunol,* 8**,** 1064-9.

Collen, D. (1998) Staphylokinase: a potent, uniquely fibrin-selective thrombolytic agent. *Nat Med,* 4**,** 279-84.

Cooke, F. J., Howard, J. C., Hugh-Jones, C. & Brown, N. M. (2008) Meticillin-resistant *Staphylococcus aureus* in the community: homeless are also at risk. *J Hosp Infect*.

Cookson, B. D., Aparicio, P., Deplano, A., Struelens, M., Goering, R. & Marples, R. (1996) Inter-centre comparison of pulsed-field gel electrophoresis for the typing of methicillin-resistant *Staphylococcus aureus*. *J Med Microbiol,* 44**,** 179-84.

Coombs, G. W., Nimmo, G. R., Bell, J. M., Huygens, F., O'brien, F. G., Malkowski, M. J., Pearson, J. C., Stephens, A. J. & Giffard, P. M. (2004) Genetic

diversity among community methicillin-resistant *Staphylococcus aureus* strains causing outpatient infections in Australia. *J Clin Microbiol,* 42**,** 4735-43.

Cosgrove, S. E., Qi, Y., Kaye, K. S., Harbarth, S., Karchmer, A. W. & Carmeli, Y. (2005) The impact of methicillin resistance in *Staphylococcus aureus* bacteremia on patient outcomes: mortality, length of stay, and hospital charges. *Infect Control Hosp Epidemiol,* 26**,** 166-74.

Cosgrove, S. E., Sakoulas, G., Perencevich, E. N., Schwaber, M. J., Karchmer, A. W. & Carmeli, Y. (2003) Comparison of mortality associated with methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* bacteremia: a meta-analysis. *Clin Infect Dis,* 36**,** 53-9.

Couto, I., De Lencastre, H., Severina, E., Kloos, W., Webster, J. A., Hubner, R. J., Sanches, I. S. & Tomasz, A. (1996) Ubiquitous presence of a *mec*A homologue in natural isolates of *Staphylococcus sciuri. Microb Drug Resist,* 2**,** 377-91.

Craven, D. E., Rixinger, A. I., Goularte, T. A. & Mccabe, W. R. (1986) Methicillin-resistant *Staphylococcus aureus* bacteremia linked to intravenous drug abusers using a "shooting gallery". *Am J Med,* 80**,** 770-6.

Creech, C. B., 2nd, Kernodle, D. S., Alsentzer, A., Wilson, C. & Edwards, K. M. (2005) Increasing rates of nasal carriage of methicillin-resistant *Staphylococcus aureus* in healthy children. *Pediatr Infect Dis J,* 24**,** 617-21.

Cui, L., Murakami, H., Kuwahara-Arai, K., Hanaki, H. & Hiramatsu, K. (2000) Contribution of a thickened cell wall and its glutamine nonamidated component to the vancomycin resistance expressed by *Staphylococcus aureus* Mu50. *Antimicrob Agents Chemother,* 44**,** 2276-85.

Cunningham, R., Cockayne, A. & Humphreys, H. (1996) Clinical and molecular aspects of the pathogenesis of *Staphylococcus aureus* bone and joint infections. *J Med Microbiol,* 44**,** 157-64.

Cunningham, R., Jenks, P., Northwood, J., Wallis, M., Ferguson, S. & Hunt, S. (2007) Effect on MRSA transmission of rapid PCR testing of patients admitted to critical care. *J Hosp Infect,* 65**,** 24-8.

Cuny, C. & Witte, W. (2005) PCR for the identification of methicillin-resistant *Staphylococcus aureus* (MRSA) strains using a single primer pair specific for SCC*mec* elements and the neighbouring chromosome-borne *orfX. Clin Microbiol Infect,* 11**,** 834-7.

Daum, R. S., Ito, T., Hiramatsu, K., Hussain, F., Mongkolrattanothai, K., Jamklang, M. & Boyle-Vavra, S. (2002) A novel methicillin-resistance cassette in community-acquired methicillin-resistant *Staphylococcus aureus* isolates of diverse genetic backgrounds. *J Infect Dis,* 186**,** 1344-7.

Davis, K. A., Stewart, J. J., Crouch, H. K., Florez, C. E. & Hospenthal, D. R. (2004) Methicillin-resistant *Staphylococcus aureus* (MRSA) nares colonization at hospital admission and its effect on subsequent MRSA infection. *Clin Infect Dis,* 39**,** 776-82.

De Haas, C. J., Veldkamp, K. E., Peschel, A., Weerkamp, F., Van Wamel, W. J., Heezius, E. C., Poppelier, M. J., Van Kessel, K. P. & Van Strijp, J. A. (2004) Chemotaxis inhibitory protein of *Staphylococcus aureus*, a bacterial antiinflammatory agent. *J Exp Med,* 199**,** 687-95.

Desai, D., Desai, N., Nightingale, P., Elliott, T. & Neuberger, J. (2003) Carriage of methicillin-resistant *Staphylococcus aureus* is associated with an increased risk of infection after liver transplantation. *Liver Transpl,* 9**,** 754-9.

Deshpande, L. M., Fritsche, T. R. & Jones, R. N. (2004) Molecular epidemiology of selected multidrug-resistant bacteria: a global report from the SENTRY Antimicrobial Surveillance Program. *Diagn Microbiol Infect Dis,* 49**,** 231-6.

Deurenberg, R. H., Vink, C., Oudhuis, G. J., Mooij, J. E., Driessen, C., Coppens, G., Craeghs, J., De Brauwer, E., Lemmen, S., Wagenvoort, H., Friedrich, A. W., Scheres, J. & Stobberingh, E. E. (2005) Different clonal complexes of methicillin-resistant *Staphylococcus aureus* are disseminated in the Euregio Meuse-Rhine region. *Antimicrob Agents Chemother,* 49**,** 4263-71.

Diekema, D. J., Pfaller, M. A., Schmitz, F. J., Smayevsky, J., Bell, J., Jones, R. N. & Beach, M. (2001) Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997-1999. *Clin Infect Dis,* 32 Suppl 2**,** S114-32.

Diep, B. A., Gill, S. R., Chang, R. F., Phan, T. H., Chen, J. H., Davidson, M. G., Lin, F., Lin, J., Carleton, H. A., Mongodin, E. F., Sensabaugh, G. F. & Perdreau-Remington, F. (2006) Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet,* 367**,** 731-9.

Diep, B. A., Perdreau-Remington, F. & Sensabaugh, G. F. (2003) Clonal characterization of *Staphylococcus aureus* by multilocus restriction fragment typing, a rapid screening approach for molecular epidemiology. *J Clin Microbiol,* 41**,** 4559-64.

Diep, B. A., Sensabaugh, G. F., Somboona, N. S., Carleton, H. A. & Perdreau-Remington, F. (2004) Widespread skin and soft-tissue infections due to two methicillin-resistant *Staphylococcus aureus* strains harboring the genes for Panton-Valentine leucocidin. *J Clin Microbiol,* 42**,** 2080-4.

Donnio, P. Y., Oliveira, D. C., Faria, N. A., Wilhelm, N., Le Coustumier, A. & De Lencastre, H. (2005) Partial excision of the chromosomal cassette containing the methicillin resistance determinant results in methicillin-susceptible *Staphylococcus aureus*. *J Clin Microbiol,* 43**,** 4191-3.

Donowitz, L. G., Wenzel, R. P. & Hoyt, J. W. (1982) High risk of hospital-acquired infection in the ICU patient. *Crit Care Med,* 10**,** 355-7.

Dubin, D. T., Matthews, P. R., Chikramane, S. G. & Stewart, P. R. (1991) Physical mapping of the *mec* region of an American methicillin-resistant *Staphylococcus aureus* strain. *Antimicrob Agents Chemother,* 35**,** 1661-5.

Dufresne, S. D., Belloni, D. R., Wells, W. A. & Tsongalis, G. J. (2006) BRCA1 and BRCA2 mutation screening using SmartCycler II high-resolution melt curve analysis. *Arch Pathol Lab Med,* 130**,** 185-7.

Enright, M. C., Day, N. P., Davies, C. E., Peacock, S. J. & Spratt, B. G. (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol,* 38**,** 1008-15.

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H. & Spratt, B. G. (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc Natl Acad Sci U S A,* 99**,** 7687-92.

Enright, M. C. & Spratt, B. G. (1999) Multilocus sequence typing. *Trends Microbiol,* 7**,** 482-7.

Eriksen, N. H., Espersen, F., Rosdahl, V. T. & Jensen, K. (1995) Carriage of *Staphylococcus aureus* among 104 healthy persons during a 19-month period. *Epidemiol Infect,* 115**,** 51-60.

Faria, N. A., Carrico, J. A., Oliveira, D. C., Ramirez, M. & De Lencastre, H. (2008) Analysis of typing methods for epidemiological surveillance of both methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains. *J Clin Microbiol,* 46**,** 136-44.

Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., Moore, C. E. & Day, N. P. (2003) How clonal is *Staphylococcus aureus*? *J Bacteriol,* 185**,** 3307-16.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol,* 186**,** 1518-30.

Fey, P. D., Said-Salim, B., Rupp, M. E., Hinrichs, S. H., Boxrud, D. J., Davis, C. C., Kreiswirth, B. N. & Schlievert, P. M. (2003) Comparative molecular analysis of community- or hospital-acquired methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 47**,** 196-203.

Fitzgerald, J. R., Reid, S. D., Ruotsalainen, E., Tripp, T. J., Liu, M., Cole, R., Kuusela, P., Schlievert, P. M., Jarvinen, A. & Musser, J. M. (2003) Genome diversification in *Staphylococcus aureus*: Molecular evolution of a highly variable chromosomal region encoding the Staphylococcal exotoxin-like family of proteins. *Infect Immun,* 71**,** 2827-38.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. (2001) Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci U S A,* 98**,** 8821-6.

Fleming, A. (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzæ*. *Br J Exp Pathol* 10**,** 226-36.

Foster, T. J. & Hook, M. (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol,* 6**,** 484-8.

Francois, P., Huyghe, A., Charbonnier, Y., Bento, M., Herzig, S., Topolski, I., Fleury, B., Lew, D., Vaudaux, P., Harbarth, S., Van Leeuwen, W., Van Belkum, A., Blanc, D. S., Pittet, D. & Schrenzel, J. (2005) Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of *Staphylococcus aureus* isolates. *J Clin Microbiol,* 43**,** 3346-55.

Francois, P., Renzi, G., Pittet, D., Bento, M., Lew, D., Harbarth, S., Vaudaux, P. & Schrenzel, J. (2004) A novel multiplex real-time PCR assay for rapid typing of major staphylococcal cassette chromosome *mec* elements. *J Clin Microbiol,* 42**,** 3309-12.

Frenay, H. M., Bunschoten, A. E., Schouls, L. M., Van Leeuwen, W. J., Vandenbroucke-Grauls, C. M., Verhoef, J. & Mooi, F. R. (1996) Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis,* 15**,** 60-4.

Gilbert, M., Macdonald, J., Gregson, D., Siushansian, J., Zhang, K., Elsayed, S., Laupland, K., Louie, T., Hope, K., Mulvey, M., Gillespie, J., Nielsen, D., Wheeler, V., Louie, M., Honish, A., Keays, G. & Conly, J. (2006) Outbreak

in Alberta of community-acquired (USA300) methicillin-resistant *Staphylococcus aureus* in people with a history of drug use, homelessness or incarceration. *Cmaj,* 175**,** 149-54.

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., Deboy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J., Khouri, H., Utterback, T., Lee, C., Dimitrov, G., Jiang, L., Qin, H., Weidman, J., Tran, K., Kang, K., Hance, I. R., Nelson, K. E. & Fraser, C. M. (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J Bacteriol,* 187**,** 2426-38.

Gillaspy, A. F., Worrell,V., Orvis,J., Roe,B.A., Dyer,D.W. And Iandolo,J.J. (2006) The *Staphylococcus aureus* NCTC8325 Genome. In Fischetti, V., Novick, R., Ferretti, J., Portnoy, D. and Rood, J. (Ed.) *Gram Positive Pathogens.* ASM Press

Greene, C., Mcdevitt, D., Francois, P., Vaudaux, P. E., Lew, D. P. & Foster, T. J. (1995) Adhesion properties of mutants of *Staphylococcus aureus* defective in fibronectin-binding proteins and studies on the expression of *fnb* genes. *Mol Microbiol,* 17**,** 1143-52.

Groom, A. V., Wolsey, D. H., Naimi, T. S., Smith, K., Johnson, S., Boxrud, D., Moore, K. A. & Cheek, J. E. (2001) Community-acquired methicillin-resistant *Staphylococcus aureus* in a rural American Indian community. *Jama,* 286**,** 1201-5.

Guinan, M. E., Dan, B. B., Guidotti, R. J., Reingold, A. L., Schmid, G. P., Bettoli, E. J., Lossick, J. G., Shands, K. N., Kramer, M. A., Hargrett, N. T., Anderson, R. L. & Broome, C. V. (1982) Vaginal colonization with *Staphylococcus aureus* in healthy women: a review of four studies. *Ann Intern Med,* 96**,** 944-7.

Gundry, C. N., Vandersteen, J. G., Reed, G. H., Pryor, R. J., Chen, J. & Wittwer, C. T. (2003) Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clin Chem,* 49**,** 396-406.

Gurtler, V., Barrie, H. D. & Mayall, B. C. (2002) Denaturing gradient gel electrophoretic multilocus sequence typing of *Staphylococcus aureus* isolates. *Electrophoresis,* 23**,** 3310-20.

Guss, B., Uhlen, M., Nilsson, B., Lindberg, M., Sjoquist, J. & Sjodahl, J. (1984) Region X, the cell-wall-attachment part of staphylococcal protein A. *Eur J Biochem,* 138**,** 413-20.

Hallin, M., Deplano, A., Denis, O., De Mendonca, R., De Ryck, R. & Struelens, M. J. (2007) Validation of pulsed-field gel electrophoresis and *spa* typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J Clin Microbiol,* 45**,** 127-33.

Hanaki, H., Kuwahara-Arai, K., Boyle-Vavra, S., Daum, R. S., Labischinski, H. & Hiramatsu, K. (1998) Activated cell-wall synthesis is associated with vancomycin resistance in methicillin-resistant *Staphylococcus aureus* clinical strains Mu3 and Mu50. *J Antimicrob Chemother,* 42**,** 199-209.

Hanssen, A. M. & Sollid, J. U. (2007) Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus*

aureus and coagulase-negative staphylococci from Norway. *Antimicrob Agents Chemother,* 51**,** 1671-7.

Harmsen, D., Claus, H. & Vogel, U. (2005) DNA sequence-based tandem repeat analysis of the *clf*B gene is less discriminatory than *spa* typing for methicillin-resistant *Staphylococcus aureus. Int J Med Microbiol,* 294**,** 525-8.

Harmsen, D., Claus, H., Witte, W., Rothganger, J., Turnwald, D. & Vogel, U. (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. *J Clin Microbiol,* 41**,** 5442-8.

Hartford, O., Francois, P., Vaudaux, P. & Foster, T. J. (1997) The dipeptide repeat region of the fibrinogen-binding protein (clumping factor) is required for functional expression of the fibrinogen-binding domain on the *Staphylococcus aureus* cell surface. *Mol Microbiol,* 25**,** 1065-76.

Hartman, B. J. & Tomasz, A. (1984) Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus. J Bacteriol,* 158**,** 513-6.

Hayes, M. V., N. A. C. Curtiss, A. W. Wyke, and J. B. Ward. (1981) Decreased affinity of a penicillin-binding protein for beta-lactam antibiotics in a clinical isolate of *Staphylococcus aureus* resistant to methicillin. *FEMS Microbiol. Lett.,* 10.

Herman, A., Kappler, J. W., Marrack, P. & Pullen, A. M. (1991) Superantigens: mechanism of T-cell stimulation and role in immune responses. *Annu Rev Immunol,* 9**,** 745-72.

Herron-Olson, L., Fitzgerald, J. R., Musser, J. M. & Kapur, V. (2007) Molecular Correlates of Host Specialization in *Staphylococcus aureus. PLoS ONE,* 2**,** e1120.

Heusser, R., Ender, M., Berger-Bachi, B. & Mccallum, N. (2007) Mosaic staphylococcal cassette chromosome *mec* containing two recombinase loci and a new *mec* complex, B2. *Antimicrob Agents Chemother,* 51**,** 390-3.

Hidron, A. I., Kourbatova, E. V., Halvosa, J. S., Terrell, B. J., Mcdougal, L. K., Tenover, F. C., Blumberg, H. M. & King, M. D. (2005) Risk factors for colonization with methicillin-resistant *Staphylococcus aureus* (MRSA) in patients admitted to an urban hospital: emergence of community-associated MRSA nasal carriage. *Clin Infect Dis,* 41**,** 159-66.

Highlander, S. K., Hulten, K. G., Qin, X., Jiang, H., Yerrapragada, S., Mason, E. O., Shang, Y., Williams, T. M., Fortunov, R. M., Liu, Y., Igboeli, O., Petrosino, J., Tirumalai, M., Uzman, A., Fox, G. E., Cardenas, A. M., Muzny, D. M., Hemphill, L., Ding, Y., Dugan, S., Blyth, P. R., Buhay, C. J., Dinh, H. H., Hawes, A. C., Holder, M., Kovar, C. L., Lee, S. L., Liu, W., Nazareth, L. V., Wang, Q., Zhou, J., Kaplan, S. L. & Weinstock, G. M. (2007) Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus. BMC Microbiol,* 7**,** 99.

Hijioka, S., Nakata, R., Yoshinami, Y., Sugita, K. & Nishimura, T. (2002) [Survey of susceptibility of methicillin-resistant *Staphylococcus aureus* to antimicrobial agents in Hokusetsu General Hospital]. *Jpn J Antibiot,* 55**,** 764-70.

Hilden, P., Savolainen, K., Tyynela, J., Vuento, M. & Kuusela, P. (1996) Purification and characterisation of a plasmin-sensitive surface protein of *Staphylococcus aureus. Eur J Biochem,* 236**,** 904-10.

Hiramatsu, K. (1998) Vancomycin resistance in staphylococci. *Drug Resist Updat,* 1**,** 135-50.

189

Hiramatsu, K. (2001) Vancomycin-resistant *Staphylococcus aureus*: a new model of antibiotic resistance. *Lancet Infect Dis,* 1, 147-55.

Hiramatsu, K., Aritaka, N., Hanaki, H., Kawasaki, S., Hosoda, Y., Hori, S., Fukuchi, Y. & Kobayashi, I. (1997a) Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin. *The Lancet,* 350, 1670-1673.

Hiramatsu, K., Asada, K., Suzuki, E., Okonogi, K. & Yokota, T. (1992) Molecular cloning and nucleotide sequence determination of the regulator region of *mec*A gene in methicillin-resistant *Staphylococcus aureus* (MRSA). *FEBS Lett,* 298, 133-6.

Hiramatsu, K. & Hanaki, H. (1998) Glycopeptide resistance in staphylococci. *Curr Opin Infect Dis,* 11, 653-8.

Hiramatsu, K., Hanaki, H., Ino, T., Yabuta, K., Oguri, T. & Tenover, F. C. (1997b) Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. *J Antimicrob Chemother,* 40, 135-6.

Hisata, K., Kuwahara-Arai, K., Yamanoto, M., Ito, T., Nakatomi, Y., Cui, L., Baba, T., Terasawa, M., Sotozono, C., Kinoshita, S., Yamashiro, Y. & Hiramatsu, K. (2005) Dissemination of methicillin-resistant staphylococci among healthy Japanese children. *J Clin Microbiol,* 43, 3364-72.

Holden, M. T., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G. & Parkhill, J. (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A,* 101, 9786-91.

Hu, L., Umeda, A., Kondo, S. & Amako, K. (1995) Typing of *Staphylococcus aureus* colonising human nasal carriers by pulsed-field gel electrophoresis. *J Med Microbiol,* 42, 127-32.

Huang, S. S., Yokoe, D. S., Hinrichsen, V. L., Spurchise, L. S., Datta, R., Miroshnik, I. & Platt, R. (2006) Impact of routine intensive care unit surveillance cultures and resultant barrier precautions on hospital-wide methicillin-resistant *Staphylococcus aureus* bacteremia. *Clin Infect Dis,* 43, 971-8.

Huang, Y. C., Hwang, K. P., Chen, P. Y., Chen, C. J. & Lin, T. Y. (2007) Prevalence of methicillin-resistant *Staphylococcus aureus* nasal colonization among Taiwanese children in 2005 and 2006. *J Clin Microbiol,* 45, 3992-5.

Huletsky, A., Giroux, R., Rossbach, V., Gagnon, M., Vaillancourt, M., Bernier, M., Gagnon, F., Truchon, K., Bastien, M., Picard, F. J., Van Belkum, A., Ouellette, M., Roy, P. H. & Bergeron, M. G. (2004) New real-time PCR assay for rapid detection of methicillin-resistant *Staphylococcus aureus* directly from specimens containing a mixture of staphylococci. *J Clin Microbiol,* 42, 1875-84.

Hunter, P. R. & Gaston, M. A. (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol,* 26, 2465-6.

Huygens, F., Inman-Bamber, J., Nimmo, G. R., Munckhof, W., Schooneveldt, J., Harrison, B., Mcmahon, J. A. & Giffard, P. M. (2006) *Staphylococcus aureus* Genotyping Using Novel Real-Time PCR Formats. *J Clin Microbiol,* 44**,** 3712-9.

Huygens, F., Nimmo, G. R., Schooneveldt, J., Munckhof, W. J. & Giffard, P. M. (2002) Genotyping of methicillin-resistant *Staphylococcus aureus* by assaying for the presence of variable elements associated with *mec*A. *J Clin Microbiol,* 40**,** 3093-7.

Huygens, F., Stephens, A. J., Nimmo, G. R. & Giffard, P. M. (2004) *mec*A Locus diversity in methicillin-resistant *Staphylococcus aureus* isolates in Brisbane, Australia, and the development of a novel diagnostic procedure for the Western Samoan phage pattern clone. *J Clin Microbiol,* 42**,** 1947-55.

Iandolo, J. J., Worrell, V., Groicher, K. H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., Qi, S., Zhu, H. & Roe, B. A. (2002) Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene,* 289**,** 109-18.

Inglis, B., Matthews, P. R. & Stewart, P. R. (1988) The expression in *Staphylococcus aureus* of cloned DNA encoding methicillin resistance. *J Gen Microbiol,* 134**,** 1465-9.

Ito, T. & Hiramatsu, K. (1998) Acquisition of methicillin resistance and progression of multiantibiotic resistance in methicillin-resistant *Staphylococcus aureus. Yonsei Med J,* 39**,** 526-33.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C. & Hiramatsu, K. (2001) Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus. Antimicrob Agents Chemother,* 45**,** 1323-36.

Ito, T., Katayama, Y. & Hiramatsu, K. (1999) Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob Agents Chemother,* 43**,** 1449-58.

Ito, T., Ma, X. X., Takeuchi, F., Okuma, K., Yuzawa, H. & Hiramatsu, K. (2004) Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccr*C. *Antimicrob Agents Chemother,* 48**,** 2637-51.

Ito, T., Okuma, K., Ma, X. X., Yuzawa, H. & Hiramatsu, K. (2003) Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: genomic island SCC. *Drug Resist Updat,* 6**,** 41-52.

Jackson, M. P. & Iandolo, J. J. (1986) Cloning and expression of the exfoliative toxin B gene from *Staphylococcus aureus. J Bacteriol,* 166**,** 574-80.

Jarraud, S., Mougel, C., Thioulouse, J., Lina, G., Meugnier, H., Forey, F., Nesme, X., Etienne, J. & Vandenesch, F. (2002) Relationships between *Staphylococcus aureus* genetic background, virulence factors, *agr* groups (alleles), and human disease. *Infect Immun,* 70**,** 631-41.

Jarraud, S., Peyrat, M. A., Lim, A., Tristan, A., Bes, M., Mougel, C., Etienne, J., Vandenesch, F., Bonneville, M. & Lina, G. (2001) *egc*, a highly prevalent operon of enterotoxin gene, forms a putative nursery of superantigens in *Staphylococcus aureus. J Immunol,* 166**,** 669-77.

Jevons, M. P. (1961) "Celbenin"- resistant *staphylococci. Br. Med. J.,* 1**,** 124-125.

Ji, G., Beavis, R. & Novick, R. P. (1997) Bacterial interference caused by autoinducing peptide variants. *Science,* 276**,** 2027-30.

Ji, G., Beavis, R. C. & Novick, R. P. (1995) Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc Natl Acad Sci U S A,* 92**,** 12055-9.

Johnson, A. P., Aucken, H. M., Cavendish, S., Ganner, M., Wale, M. C., Warner, M., Livermore, D. M. & Cookson, B. D. (2001) Dominance of EMRSA-15 and - 16 among MRSA causing nosocomial bacteraemia in the UK: analysis of isolates from the European Antimicrobial Resistance Surveillance System (EARSS). *J Antimicrob Chemother,* 48**,** 143-4.

Josefsson, E., M*ccr*ea, K. W., Ni Eidhin, D., O'connell, D., Cox, J., Hook, M. & Foster, T. J. (1998) Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus. Microbiology,* 144 (Pt 12)**,** 3387-95.

Juuti, K., Ibrahem, S., Virolainen-Julkunen, A., Vuopio-Varkila, J. & Kuusela, P. (2005) The *pls* gene found in methicillin-resistant *Staphylococcus aureus* strains is common in clinical isolates of *Staphylococcus sciuri. J Clin Microbiol,* 43**,** 1415-9.

Katayama, Y., Ito, T. & Hiramatsu, K. (2000) A new class of genetic element, *staphylococcus* cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus. Antimicrob Agents Chemother,* 44**,** 1549-55.

Katayama, Y., Ito, T. & Hiramatsu, K. (2001) Genetic organization of the chromosome region surrounding *mec*A in clinical staphylococcal strains: role of IS*431*-mediated *mec*I deletion in expression of resistance in *mec*A-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus. Antimicrob Agents Chemother,* 45**,** 1955-63.

Katayama, Y., Takeuchi, F., Ito, T., Ma, X. X., Ui-Mizutani, Y., Kobayashi, I. & Hiramatsu, K. (2003) Identification in methicillin-susceptible *Staphylococcus hominis* of an active primordial mobile genetic element for the staphylococcal cassette chromosome *mec* of methicillin-resistant *Staphylococcus aureus. J Bacteriol,* 185**,** 2711-22.

Keim, P., Price, L. B., Klevytska, A. M., Smith, K. L., Schupp, J. M., Okinaka, R., Jackson, P. J. & Hugh-Jones, M. E. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis. J Bacteriol,* 182**,** 2928-36.

Keim, P., Van Ert, M. N., Pearson, T., Vogler, A. J., Huynh, L. Y. & Wagner, D. M. (2004) Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect Genet Evol,* 4**,** 205-13.

Khanna, T., Friendship, R., Dewey, C. & Weese, J. S. (2008) Methicillin resistant *Staphylococcus aureus* colonization in pigs and pig farmers. *Vet Microbiol,* 128**,** 298-303.

Klevens, R. M., Edwards, J. R., Tenover, F. C., Mcdonald, L. C., Horan, T. & Gaynes, R. (2006) Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in intensive care units in US hospitals, 1992-2003. *Clin Infect Dis,* 42**,** 389-91.

Kluytmans, J., Van Belkum, A. & Verbrugh, H. (1997) Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev,* 10**,** 505-20.

Kluytmans, J. A. & Wertheim, H. F. (2005) Nasal carriage of *Staphylococcus aureus* and prevention of nosocomial infections. *Infection,* 33**,** 3-8.

Kobayashi, N., Alam, M. M. & Urasawa, S. (2001) Genomic rearrangement of the *mec* regulator region mediated by insertion of IS431 in methicillin-resistant staphylococci. *Antimicrob Agents Chemother,* 45**,** 335-8.

Kondo, Y., Ito, T., Ma, X. X., Watanabe, S., Kreiswirth, B. N., Etienne, J. & Hiramatsu, K. (2007) Combination of Multiplex PCRs for Staphylococcal Cassette Chromosome *mec* Type Assignment: Rapid Identification System for *mec*, *ccr*, and Major Differences in Junkyard Regions. *Antimicrob Agents Chemother,* 51**,** 264-74.

Kong, F., Ma, L. & Gilbert, G. L. (2005) Simultaneous detection and serotype identification of *Streptococcus agalactiae* using multiplex PCR and reverse line blot hybridization. *J Med Microbiol,* 54**,** 1133-8.

Koreen, L., Ramaswamy, S. V., Graviss, E. A., Naidich, S., Musser, J. M. & Kreiswirth, B. N. (2004) *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol,* 42**,** 792-9.

Koreen, L., Ramaswamy, S. V., Naidich, S., Koreen, I. V., Graff, G. R., Graviss, E. A. & Kreiswirth, B. N. (2005) Comparative sequencing of the serine-aspartate repeat-encoding region of the clumping factor B gene (*clfB*) for resolution within clonal groups of *Staphylococcus aureus*. *J Clin Microbiol,* 43**,** 3985-94.

Krypuy, M., Newnham, G. M., Thomas, D. M., Conron, M. & Dobrovic, A. (2006) High resolution melting analysis for the rapid and sensitive detection of mutations in clinical samples: KRAS codon 12 and 13 mutations in non-small cell lung cancer. *BMC Cancer,* 6**,** 295.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. & Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet,* 357**,** 1225-40.

Kuwahara-Arai, K., Kondo, N., Hori, S., Tateda-Suzuki, E. & Hiramatsu, K. (1996) Suppression of methicillin resistance in a *mec*A-containing pre-methicillin-resistant *Staphylococcus aureus* strain is caused by the *mec*I-mediated repression of PBP 2' production. *Antimicrob Agents Chemother,* 40**,** 2680-5.

Kwon, N. H., Park, K. T., Moon, J. S., Jung, W. K., Kim, S. H., Kim, J. M., Hong, S. K., Koo, H. C., Joo, Y. S. & Park, Y. H. (2005) Staphylococcal cassette chromosome *mec* (SCC*mec*) characterization and molecular analysis for methicillin-resistant *Staphylococcus aureus* and novel SCC*mec* subtype IVg isolated from bovine milk in Korea. *J Antimicrob Chemother,* 56**,** 624-32.

Labandeira-Rey, M., Couzon, F., Boisset, S., Brown, E. L., Bes, M., Benito, Y., Barbu, E. M., Vazquez, V., Hook, M., Etienne, J., Vandenesch, F. & Bowden, M. G. (2007) *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. *Science,* 315**,** 1130-3.

Ladhani, S., Joannou, C. L., Lochrie, D. P., Evans, R. W. & Poston, S. M. (1999) Clinical, microbial, and biochemical aspects of the exfoliative toxins causing staphylococcal scalded-skin syndrome. *Clin Microbiol Rev,* 12**,** 224-42.

Layer, F., Ghebremedhin, B., Konig, W. & Konig, B. (2006) Heterogeneity of methicillin-susceptible *Staphylococcus aureus* strains at a German University

Hospital implicates the circulating-strain pool as a potential source of emerging methicillin-resistant S. aureus clones. *J Clin Microbiol,* 44, 2179-85.

Le Loir, Y., Baron, F. & Gautier, M. (2003) *Staphylococcus aureus* and food poisoning. *Genet Mol Res,* 2, 63-76.

Lee, N. E., Taylor, M. M., Bancroft, E., Ruane, P. J., Morgan, M., Mccoy, L. & Simon, P. A. (2005) Risk factors for community-associated methicillin-resistant *Staphylococcus aureus* skin infections among HIV-positive men who have sex with men. *Clin Infect Dis,* 40, 1529-34.

Lim, T. T., Chong, F. N., O'brien, F. G. & Grubb, W. B. (2003) Are all community methicillin-resistant *Staphylococcus aureus* related? A comparison of their *mec* regions. *Pathology,* 35, 336-43.

Lim, T. T., Coombs, G. W. & Grubb, W. B. (2002) Genetic organization of *mec*A and *mec*A-regulatory genes in epidemic methicillin-resistant *Staphylococcus aureus* from Australia and England. *J Antimicrob Chemother,* 50, 819-24.

Lin, Y. C., Lauderdale, T. L., Lin, H. M., Chen, P. C., Cheng, M. F., Hsieh, K. S. & Liu, Y. C. (2007) An outbreak of methicillin-resistant *Staphylococcus aureus* infection in patients of a pediatric intensive care unit and high carriage rate among health care workers. *J Microbiol Immunol Infect,* 40, 325-34.

Lina, G., Boutite, F., Tristan, A., Bes, M., Etienne, J. & Vandenesch, F. (2003) Bacterial competition for human nasal cavity colonization: role of Staphylococcal *agr* alleles. *Appl Environ Microbiol,* 69, 18-23.

Lina, G., Durand, G., Berchich, C., Short, B., Meugnier, H., Vandenesch, F., Etienne, J. & Enright, M. C. (2006) Staphylococcal chromosome cassette evolution in *Staphylococcus aureus* inferred from *ccr* gene complex sequence typing analysis. *Clin Microbiol Infect,* 12, 1175-84.

Lina, G., Jarraud, S., Ji, G., Greenland, T., Pedraza, A., Etienne, J., Novick, R. P. & Vandenesch, F. (1998) Transmembrane topology and histidine protein kinase activity of *Agr*C, the *agr* signal receptor in *Staphylococcus aureus. Mol Microbiol,* 28, 655-62.

Lindenmayer, J. M., Schoenfeld, S., O'grady, R. & Carney, J. K. (1998) Methicillin-resistant *Staphylococcus aureus* in a high school wrestling team and the surrounding community. *Arch Intern Med,* 158, 895-9.

Lindsay, J. A. & Holden, M. T. (2004) *Staphylococcus aureus*: superbug, super genome? *Trends Microbiol,* 12, 378-85.

Lindsay, J. A. & Holden, M. T. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus. Funct Integr Genomics,* 6, 186-201.

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D. & Hinds, J. (2006) Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. *J Bacteriol,* 188, 669-76.

Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus. Mol Microbiol,* 29, 527-43.

Livak, K. J. (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal,* 14, 143-9.

Lofdahl, S., Guss, B., Uhlen, M., Philipson, L. & Lindberg, M. (1983) Gene for staphylococcal protein A. *Proc Natl Acad Sci U S A,* 80, 697-701.

Ma, X. X., Ito, T., Tiensasitorn, C., Jamklang, M., Chongtrakool, P., Boyle-Vavra, S., Daum, R. S. & Hiramatsu, K. (2002) Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. *Antimicrob Agents Chemother,* 46, 1147-52.

Madigan, M., Martinko, J. & Parker, J. (2000a) Microbial Genetics. *Biology of Microorganisms.* Ninth ed. New Jersey, Prentice-Hall, Inc.

Madigan, M., Martinko, J. & Parker, J. (2000b) Prokaryote Diversity: Bacteria. *Biology of Microorganisms.* Ninth ed. New Jersey, Prentice-Hall, Inc.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A,* 95, 3140-5.

Malachowa, N., Sabat, A., Gniadkowski, M., Krzyszton-Russjan, J., Empel, J., Miedzobrodzki, J., Kosowska-Shick, K., Appelbaum, P. C. & Hryniewicz, W. (2005) Comparison of multiple-locus variable-number tandem-repeat analysis with pulsed-field gel electrophoresis, *spa* typing, and multilocus sequence typing for clonal characterization of *Staphylococcus aureus* isolates. *J Clin Microbiol,* 43, 3095-100.

Maslow, J. N., Mulligan, M. E. & Arbeit, R. D. (1993) Molecular epidemiology: application of contemporary techniques to the typing of microorganisms. *Clin Infect Dis,* 17, 153-62; quiz 163-4.

Matsuhashi, M., Song, M. D., Ishino, F., Wachi, M., Doi, M., Inoue, M., Ubukata, K., Yamashita, N. & Konno, M. (1986) Molecular cloning of the gene of a penicillin-binding protein supposed to cause high resistance to beta-lactam antibiotics in *Staphylococcus aureus. J Bacteriol,* 167, 975-80.

Mazmanian, S. K., Ton-That, H. & Schneewind, O. (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus. Mol Microbiol,* 40, 1049-57.

Mcdevitt, D., Francois, P., Vaudaux, P. & Foster, T. J. (1994) Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus. Mol Microbiol,* 11, 237-48.

Mcdougal, L. K., Steward, C. D., Killgore, G. E., Chaitram, J. M., Mcallister, S. K. & Tenover, F. C. (2003) Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. *J Clin Microbiol,* 41, 5113-20.

Mcdowell, P., Affas, Z., Reynolds, C., Holden, M. T., Wood, S. J., Saint, S., Cockayne, A., Hill, P. J., Dodd, C. E., Bycroft, B. W., Chan, W. C. & Williams, P. (2001) Structure, activity and evolution of the group I thiolactone peptide quorum-sensing system of *Staphylococcus aureus. Mol Microbiol,* 41, 503-12.

Mckenzie, T., Hoshino, T., Tanaka, T. & Sueoka, N. (1986) The nucleotide sequence of pUB110: some salient features in relation to replication and its regulation. *Plasmid,* 15, 93-103.

Melles, D. C., Tenover, F. C., Kuehnert, M. J., Witsenboer, H., Peeters, J. K., Verbrugh, H. A. & Van Belkum, A. (2008) Overlapping population structures of nasal isolates of *Staphylococcus aureus* from healthy Dutch and American individuals. *J Clin Microbiol,* 46, 235-41.

Melles, D. C., Van Leeuwen, W. B., Snijders, S. V., Horst-Kreft, D., Peeters, J. K., Verbrugh, H. A. & Van Belkum, A. (2007) Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. *J Microbiol Methods,* 69**,** 371-5.

Mellmann, A., Weniger, T., Berssenbrugge, C., Rothganger, J., Sammeth, M., Stoye, J. & Harmsen, D. (2007) Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of *Staphylococcus aureus* populations based on *spa* polymorphisms. *BMC Microbiol,* 7**,** 98.

Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res,* 35**,** e97.

Milheirico, C., Oliveira, D. C. & De Lencastre, H. (2007a) Multiplex PCR strategy for subtyping the staphylococcal cassette chromosome *mec* type IV in methicillin-resistant *Staphylococcus aureus*: 'SCC*mec* IV multiplex'. *J Antimicrob Chemother,* 60**,** 42-8.

Milheirico, C., Oliveira, D. C. & De Lencastre, H. (2007b) Update to the multiplex PCR strategy for assignment of *mec* element types in *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 51**,** 3374-7.

Mongkolrattanothai, K., Boyle, S., Murphy, T. V. & Daum, R. S. (2004) Novel non-*mec*A-containing staphylococcal chromosomal cassette composite island containing *pbp4* and *tagF* genes in a commensal staphylococcal species: a possible reservoir for antibiotic resistance islands in *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 48**,** 1823-36.

Montesinos, I., Salido, E., Delgado, T., Cuervo, M. & Sierra, A. (2002) Epidemiologic genotyping of methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis at a university hospital and comparison with antibiotyping and protein A and coagulase gene polymorphisms. *J Clin Microbiol,* 40**,** 2119-25.

Moore, P. C. & Lindsay, J. A. (2001) Genetic variation among hospital isolates of methicillin-sensitive *Staphylococcus aureus*: evidence for horizontal transfer of virulence genes. *J Clin Microbiol,* 39**,** 2760-7.

Munckhof, W. J., Schooneveldt, J., Coombs, G. W., Hoare, J. & Nimmo, G. R. (2003) Emergence of community-acquired methicillin-resistant *Staphylococcus aureus* (MRSA) infection in Queensland, Australia. *Int J Infect Dis,* 7**,** 259-64.

Mwangi, M. M., Wu, S. W., Zhou, Y., Sieradzki, K., De Lencastre, H., Richardson, P., Bruce, D., Rubin, E., Myers, E., Siggia, E. D. & Tomasz, A. (2007) Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Natl Acad Sci U S A,* 104**,** 9451-6.

Nakamura, Y., Leppert, M., O'connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. & Et Al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science,* 235**,** 1616-22.

Navarre, W. W. & Schneewind, O. (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol Microbiol,* 14**,** 115-21.

NCBI (2008)
http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genomeprj&cmd=ShowDetail
View&TermToSearch=19993.

Neoh, H. M., Cui, L., Yuzawa, H., Takeuchi, F., Matsuo, M. & Hiramatsu, K. (2008)
Mutated Response Regulator *graR* Is Responsible for Phenotypic Conversion
of *Staphylococcus aureus* from Heterogeneous Vancomycin-Intermediate
Resistance to Vancomycin-Intermediate Resistance. *Antimicrob Agents
Chemother,* 52**,** 45-53.

Neuzil, P., Pipper, J. & Hsieh, T. M. (2006) Disposable real-time microPCR device:
lab-on-a-chip at a low cost. *Mol Biosyst,* 2**,** 292-8.

Ni Eidhin, D., Perkins, S., Francois, P., Vaudaux, P., Hook, M. & Foster, T. J. (1998)
Clumping factor B (*Clf*B), a new surface-located fibrinogen-binding adhesin
of *Staphylococcus aureus. Mol Microbiol,* 30**,** 245-57.

Nimmo, G. R., Bell, J. M., Mitchell, D., Gosbell, I. B., Pearman, J. W. & Turnidge,
J. D. (2003) Antimicrobial resistance in *Staphylococcus aureus* in Australian
teaching hospitals, 1989-1999. *Microb Drug Resist,* 9**,** 155-60.

Nimmo, G. R., Schooneveldt, J., O'kane, G., Mccall, B. & Vickery, A. (2000)
Community acquisition of gentamicin-sensitive methicillin-resistant
Staphylococcus aureus in southeast Queensland, Australia. *J Clin Microbiol,*
38**,** 3926-31.

Noller, A. C., Mcellistrem, M. C., Pacheco, A. G., Boxrud, D. J. & Harrison, L. H.
(2003) Multilocus variable-number tandem repeat analysis distinguishes
outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J Clin Microbiol,*
41**,** 5389-97.

Nouwen, J. L., Fieren, M. W., Snijders, S., Verbrugh, H. A. & Van Belkum, A.
(2005) Persistent (not intermittent) nasal carriage of *Staphylococcus aureus* is
the determinant of CPD-related infections. *Kidney Int,* 67**,** 1084-92.

Nouwen, J. L., Ott, A., Kluytmans-Vandenbergh, M. F., Boelens, H. A., Hofman, A.,
Van Belkum, A. & Verbrugh, H. A. (2004) Predicting the *Staphylococcus
aureus* nasal carrier state: derivation and validation of a "culture rule". *Clin
Infect Dis,* 39**,** 806-11.

Novick, R. P., Adler, G. K., Majumder, S., Khan, S. A., Carleton, S., Rosenblum, W.
D. & Iordanescu, S. (1982) Coding sequence for the pT181 *repC* product: a
plasmid-coded protein uniquely required for replication. *Proc Natl Acad Sci
U S A,* 79**,** 4108-12.

Novick, R. P., Projan, S. J., Kornblum, J., Ross, H. F., Ji, G., Kreiswirth, B.,
Vandenesch, F. & Moghazeh, S. (1995) The *agr* P2 operon: an autocatalytic
sensory transduction system in *Staphylococcus aureus*. *Mol Gen Genet,* 248**,**
446-58.

Novick, R. P. & Subedi, A. (2007) The SaPIs: mobile pathogenicity islands of
Staphylococcus. Chem Immunol Allergy, 93**,** 42-57.

O'brien, F. G., Coombs, G. W., Pearson, J. C., Christiansen, K. J. & Grubb, W. B.
(2005) Type V staphylococcal cassette chromosome *mec* in community
staphylococci from Australia. *Antimicrob Agents Chemother,* 49**,** 5129-32.

Ofner-Agostini, M., Simor, A. E., Mulvey, M., Bryce, E., Loeb, M., Mcgeer, A.,
Kiss, A. & Paton, S. (2006) Methicillin-resistant *Staphylococcus aureus* in
Canadian aboriginal people. *Infect Control Hosp Epidemiol,* 27**,** 204-7.

Ogston, A. (1881) Report upon micro-organisms in surgical diseases. *BMJ,* 1**,** 369-
75.

Okuma, K., Iwakawa, K., Turnidge, J. D., Grubb, W. B., Bell, J. M., O'brien, F. G., Coombs, G. W., Pearman, J. W., Tenover, F. C., Kapi, M., Tiensasitorn, C., Ito, T. & Hiramatsu, K. (2002) Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. *J Clin Microbiol,* 40**,** 4289-94.

Oliveira, D. C. & De Lencastre, H. (2002) Multiplex PCR strategy for rapid identification of structural types and variants of the *mec* element in methicillin-resistant *Staphylococcus aureus. Antimicrob Agents Chemother,* 46**,** 2155-61.

Oliveira, D. C., Milheirico, C. & De Lencastre, H. (2006a) Redefining a structural variant of staphylococcal cassette chromosome *mec*, SCC*mec* type VI. *Antimicrob Agents Chemother,* 50**,** 3457-9.

Oliveira, D. C., Milheirico, C., Vinga, S. & De Lencastre, H. (2006b) Assessment of allelic variation in the *ccr*AB locus in methicillin-resistant *Staphylococcus aureus* clones. *J Antimicrob Chemother,* 58**,** 23-30.

Oliveira, D. C., Tomasz, A. & De Lencastre, H. (2001) The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated *mec* elements. *Microb Drug Resist,* 7**,** 349-61.

Oliveira, D. C., Wu, S. W. & De Lencastre, H. (2000) Genetic organization of the downstream region of the *mec*A element in methicillin-resistant *Staphylococcus aureus* isolates carrying different polymorphisms of this region. *Antimicrob Agents Chemother,* 44**,** 1906-10.

Otter, J. A., Klein, J. L., Watts, T. L., Kearns, A. M. & French, G. L. (2007) Identification and control of an outbreak of ciprofloxacin-susceptible EMRSA-15 on a neonatal unit. *J Hosp Infect,* 67**,** 232-9.

Panlilio, A. L., Culver, D. H., Gaynes, R. P., Banerjee, S., Henderson, T. S., Tolson, J. S. & Martone, W. J. (1992) Methicillin-resistant *Staphylococcus aureus* in U.S. hospitals, 1975- 1991. *Infect Control Hosp Epidemiol,* 13**,** 582-6.

Papageorgiou, A. C. & Acharya, K. R. (2000) Microbial superantigens: from structure to function. *Trends Microbiol,* 8**,** 369-75.

Patti, J. M. & Hook, M. (1994) Microbial adhesins recognizing extracellular matrix macromolecules. *Curr Opin Cell Biol,* 6**,** 752-8.

Patti, J. M., Jonsson, H., Guss, B., Switalski, L. M., Wiberg, K., Lindberg, M. & Hook, M. (1992) Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesin. *J Biol Chem,* 267**,** 4766-72.

Paulsen, I. T. F., N. Skurray, R.A. (1997) Resistance to antimicrobial agents other than β-lactams. IN CROSSLEY, K. B. A., G.L. (Ed.) *The Staphylococci in human disease.* New York, Churchill Livingstone.

Peacock, S. J., Moore, C. E., Justice, A., Kantzanou, M., Story, L., Mackie, K., O'neill, G. & Day, N. P. (2002) Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus. Infect Immun,* 70**,** 4987-96.

Peleg, A. Y., Munckhof, W. J., Kleinschmidt, S. L., Stephens, A. J. & Huygens, F. (2005) Life-threatening community-acquired methicillin-resistant *Staphylococcus aureus* infection in Australia. *Eur J Clin Microbiol Infect Dis,* 24**,** 384-7.

Phillips, S. & Novick, R. P. (1979) Tn*554*--a site-specific repressor-controlled transposon in *Staphylococcus aureus. Nature,* 278**,** 476-8.

Ploy, M. C., Grelaud, C., Martin, C., De Lumley, L. & Denis, F. (1998) First clinical isolate of vancomycin-intermediate *Staphylococcus aureus* in a French hospital. *Lancet,* 351**,** 1212.

Price, E. P., Huygens, F. & Giffard, P. M. (2006) Fingerprinting of *Campylobacter jejuni* by Using Resolution-Optimized Binary Gene Targets Derived from Comparative Genome Hybridization Studies. *Appl Environ Microbiol,* 72**,** 7793-803.

Price, E. P., Smith, H., Huygens, F. & Giffard, P. M. (2007) High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni. Appl Environ Microbiol,* 73**,** 3431-6.

Ridley, M. (1959) Perineal carriage of *Staph. aureus. Br Med J,* 1**,** 270-3.

Rimland, D. & Roberson, B. (1986) Gastrointestinal carriage of methicillin-resistant *Staphylococcus aureus. J Clin Microbiol,* 24**,** 137-8.

Ririe, K. M., Rasmussen, R. P. & Wittwer, C. T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem,* 245**,** 154-60.

Roberts, R. B., De Lencastre, A., Eisner, W., Severina, E. P., Shopsin, B., Kreiswirth, B. N. & Tomasz, A. (1998) Molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in 12 New York hospitals. MRSA Collaborative Study Group. *J Infect Dis,* 178**,** 164-71.

Robertson, G. A., Thiruvenkataswamy, V., Shilling, H., Price, E. P., Huygens, F., Henskens, F. A. & Giffard, P. M. (2004) Identification and interrogation of highly informative single nucleotide polymorphism sets defined by bacterial multilocus sequence typing databases. *J Med Microbiol,* 53**,** 35-45.

Robinson, D. A. & Enright, M. C. (2003) Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus. Antimicrob Agents Chemother,* 47**,** 3926-34.

Robinson, D. A. & Enright, M. C. (2004a) Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol,* 186**,** 1060-4.

Robinson, D. A. & Enright, M. C. (2004b) Multilocus sequence typing and the evolution of methicillin-resistant *Staphylococcus aureus. Clin Microbiol Infect,* 10**,** 92-7.

Robinson, D. A., Kearns, A. M., Holmes, A., Morrison, D., Grundmann, H., Edwards, G., O'brien, F. G., Tenover, F. C., Mcdougal, L. K., Monk, A. B. & Enright, M. C. (2005) Re-emergence of early pandemic *Staphylococcus aureus* as a community-acquired meticillin-resistant clone. *Lancet,* 365**,** 1256-8.

Rolinson, G. N., Stevens, S., Batchelor, F. R., Wood, J. C. & Chain, E. B. (1960) Bacteriological studies on a new penicillin-BRL. 1241. *Lancet,* 2**,** 564-7.

Rosdahl, V. T. & Knudsen, A. M. (1991) The decline of methicillin resistance among Danish *Staphylococcus aureus* strains. *Infect Control Hosp Epidemiol,* 12**,** 83-8.

Ruzin, A., Lindsay, J. & Novick, R. P. (2001) Molecular genetics of SaPI1--a mobile pathogenicity island in *Staphylococcus aureus. Mol Microbiol,* 41**,** 365-77.

Ryffel, C., Kayser, F. H. & Berger-Bachi, B. (1992) Correlation between regulation of *mec*A transcription and expression of methicillin resistance in staphylococci. *Antimicrob Agents Chemother,* 36**,** 25-31.

Sabat, A., Krzyszton-Russjan, J., Strzalka, W., Filipek, R., Kosowska, K., Hryniewicz, W., Travis, J. & Potempa, J. (2003) New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat

analysis of polymorphism and genetic relationships of clinical isolates. *J Clin Microbiol,* 41**,** 1801-4.

Salgado, C. D., Farr, B. M. & Calfee, D. P. (2003) Community-acquired methicillin-resistant *Staphylococcus aureus*: a meta-analysis of prevalence and risk factors. *Clin Infect Dis,* 36**,** 131-9.

Sanger Institute, W. T. http://www.sanger.ac.uk/sequencing/Staphylococcus/aureus/0582/.

Sanger Institute, W. T. http://www.sanger.ac.uk/sequencing/Staphylococcus/aureus/EMRSA15/.

Schnell, N., Entian, K. D., Schneider, U., Gotz, F., Zahner, H., Kellner, R. & Jung, G. (1988) Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature,* 333**,** 276-8.

Schwartz, D. C. & Cantor, C. R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell,* 37**,** 67-75.

Shopsin, B., Gomez, M., Montgomery, S. O., Smith, D. H., Waddington, M., Dodge, D. E., Bost, D. A., Riehman, M., Naidich, S. & Kreiswirth, B. N. (1999) Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J Clin Microbiol,* 37**,** 3556-63.

Shore, A., Rossney, A. S., Keane, C. T., Enright, M. C. & Coleman, D. C. (2005) Seven novel variants of the staphylococcal chromosomal cassette *mec* in methicillin-resistant *Staphylococcus aureus* isolates from Ireland. *Antimicrob Agents Chemother,* 49**,** 2070-83.

Shukla, S. K., Ramaswamy, S. V., Conradt, J., Stemper, M. E., Reich, R., Reed, K. D. & Graviss, E. A. (2004a) Novel polymorphisms in *mec* genes and a new *mec* complex type in methicillin-resistant *Staphylococcus aureus* isolates obtained in rural Wisconsin. *Antimicrob Agents Chemother,* 48**,** 3080-5.

Shukla, S. K., Stemper, M. E., Ramaswamy, S. V., Conradt, J. M., Reich, R., Graviss, E. A. & Reed, K. D. (2004b) Molecular characteristics of nosocomial and Native American community-associated methicillin-resistant *Staphylococcus aureus* clones from rural Wisconsin. *J Clin Microbiol,* 42**,** 3752-7.

Sista, R. R., Oda, G. & Barr, J. (2004) Methicillin-resistant *Staphylococcus aureus* infections in ICU patients. *Anesthesiol Clin North America,* 22**,** 405-35, vi.

Sjostrom, J. E., Lofdahl, S. & Philipson, L. (1975) Transformation reveals a chromosomal locus of the gene(s) for methicillin resistance in *Staphylococcus aureus. J Bacteriol,* 123**,** 905-15.

Smeltzer, M. S., Hart, M. E. & Iandolo, J. J. (1994) The effect of lysogeny on the genomic organization of *Staphylococcus aureus. Gene,* 138**,** 51-7.

Smith, C. L., Econome, J. G., Schutt, A., Klco, S. & Cantor, C. R. (1987) A physical map of the *Escherichia coli* K12 genome. *Science,* 236**,** 1448-53.

Smith, K. J., Wagner, K. F., Yeager, J., Skelton, H. G. & Ledsky, R. (1994) *Staphylococcus aureus* carriage and HIV-1 disease: association with increased mucocutaneous infections as well as deep soft-tissue infections and sepsis. *Arch Dermatol,* 130**,** 521-2.

Smith, T. L., Pearson, M. L., Wilcox, K. R., Cruz, C., Lancaster, M. V., Robinson-Dunn, B., Tenover, F. C., Zervos, M. J., Band, J. D., White, E. & Jarvis, W. R. (1999) Emergence of vancomycin resistance in *Staphylococcus aureus*. Glycopeptide-Intermediate *Staphylococcus aureus* Working Group. *N Engl J Med,* 340**,** 493-501.

Solberg, C. O. (2000) Spread of *Staphylococcus aureus* in hospitals: causes and prevention. *Scand J Infect Dis,* 32**,** 587-95.

Song, M. D., Wachi, M., Doi, M., Ishino, F. & Matsuhashi, M. (1987) Evolution of an inducible penicillin-target protein in methicillin-resistant *Staphylococcus aureus* by gene fusion. *FEBS Lett,* 221**,** 167-71.

Spink, W. W. & Ferris, V. (1945) Quantitative Action of Penicillin Inhibitor from Penicillin-Resistant Strains of Staphylococci. *Science,* 102**,** 221-223.

Stacey, A. R., Endersby, K. E., Chan, P. C. & Marples, R. R. (1998) An outbreak of methicillin resistant *Staphylococcus aureus* infection in a rugby football team. *Br J Sports Med,* 32**,** 153-4.

Stewart, G. C. & Rosenblum, E. D. (1980) Genetic behavior of the methicillin resistance determinant in *Staphylococcus aureus*. *J Bacteriol,* 144**,** 1200-2.

Storch, G. A., Radcliff, J. L., Meyer, P. L. & Hinrichs, J. H. (1987) Methicillin-resistant *Staphylococcus aureus* in a nursing home. *Infect Control,* 8**,** 24-9.

Strommenger, B., Kettlitz, C., Weniger, T., Harmsen, D., Friedrich, A. W. & Witte, W. (2006) Assignment of *Staphylococcus* isolates to groups by *spa* typing, SmaI macrorestriction analysis, and multilocus sequence typing. *J Clin Microbiol,* 44**,** 2533-40.

Sugiura, A., Nakashima, K., Tanaka, K. & Mizuno, T. (1992) Clarification of the structural and functional features of the osmoregulated kdp operon of *Escherichia coli*. *Mol Microbiol,* 6**,** 1769-76.

Suzuki, E., Kuwahara-Arai, K., Richardson, J. F. & Hiramatsu, K. (1993) Distribution of *mec* regulator genes in methicillin-resistant *Staphylococcus* clinical strains. *Antimicrob Agents Chemother,* 37**,** 1219-26.

Takano, T., Saito, K., Teng, L. J. & Yamamoto, T. (2007) Spread of community-acquired methicillin-resistant *Staphylococcus aureus* (MRSA) in hospitals in Taipei, Taiwan in 2005, and comparison of its drug resistance with previous hospital-acquired MRSA. *Microbiol Immunol,* 51**,** 627-32.

Taneike, I., Otsuka, T., Dohmae, S., Saito, K., Ozaki, K., Takano, M., Higuchi, W., Takano, T. & Yamamoto, T. (2006) Molecular nature of methicillin-resistant *Staphylococcus aureus* derived from explosive nosocomial outbreaks of the 1980s in Japan. *FEBS Lett,* 580**,** 2323-34.

Tenover, F. C., Arbeit, R., Archer, G., Biddle, J., Byrne, S., Goering, R., Hancock, G., Hebert, G. A., Hill, B., Hollis, R. & Et Al. (1994) Comparison of traditional and molecular methods of typing isolates of *Staphylococcus aureus*. *J Clin Microbiol,* 32**,** 407-15.

Tenover, F. C., Arbeit, R. D. & Goering, R. V. (1997) How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol,* 18**,** 426-39.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol,* 33**,** 2233-9.

Tenover, F. C., Vaughn, R. R., Mcdougal, L. K., Fosheim, G. E. & Mcgowan, J. E., Jr. (2007) Multiple-locus variable-number tandem-repeat assay analysis of methicillin-resistant *Staphylococcus aureus* strains. *J Clin Microbiol,* 45**,** 2215-9.

Tesch, W., Ryffel, C., Strassle, A., Kayser, F. H. & Berger-Bachi, B. (1990) Evidence of a novel staphylococcal *mec*-encoded element (*mec*R) controlling expression of penicillin-binding protein 2'. *Antimicrob Agents Chemother,* 34**,** 1703-6.

Thompson, R. L., Cabezudo, I. & Wenzel, R. P. (1982) Epidemiology of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. *Ann Intern Med,* 97**,** 309-17.

Tiemersma, E. W., Bronzwaer, S. L., Lyytikainen, O., Degener, J. E., Schrijnemakers, P., Bruinsma, N., Monen, J., Witte, W. & Grundman, H. (2004) Methicillin-resistant *Staphylococcus aureus* in Europe, 1999-2002. *Emerg Infect Dis,* 10**,** 1627-34.

Tung, H., Guss, B., Hellman, U., Persson, L., Rubin, K. & Ryden, C. (2000) A bone sialoprotein-binding protein from *Staphylococcus aureus*: a member of the staphylococcal Sdr family. *Biochem J,* 345 Pt 3**,** 611-9.

Uhlen, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L. & Lindberg, M. (1984) Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. *J Biol Chem,* 259**,** 1695-702.

Van Belkum, A., Scherer, S., Van Alphen, L. & Verbrugh, H. (1998a) Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev,* 62**,** 275-93.

Van Belkum, A., Van Leeuwen, W., Kaufmann, M. E., Cookson, B., Forey, F., Etienne, J., Goering, R., Tenover, F., Steward, C., O'brien, F., Grubb, W., Tassios, P., Legakis, N., Morvan, A., El Solh, N., De Ryck, R., Struelens, M., Salmenlinna, S., Vuopio-Varkila, J., Kooistra, M., Talens, A., Witte, W. & Verbrugh, H. (1998b) Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. *J Clin Microbiol,* 36**,** 1653-9.

Van Den Braak, N., Simons, G., Gorkink, R., Reijans, M., Eadie, K., Kremers, K., Van Soolingen, D., Savelkoul, P., Verbrugh, H. & Van Belkum, A. (2004) A new high-throughput AFLP approach for identification of new genetic polymorphism in the genome of the clonal microorganism *Mycobacterium tuberculosis*. *J Microbiol Methods,* 56**,** 49-62.

Van Leeuwen, W. B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H. A., Enright, M. C., Troesch, A. & Van Belkum, A. (2003) Multilocus sequence typing of *Staphylococcus aureus* with DNA array technology. *J Clin Microbiol,* 41**,** 3323-6.

Van Trijp, M. J., Melles, D. C., Hendriks, W. D., Parlevliet, G. A., Gommans, M. & Ott, A. (2007) Successful control of widespread methicillin-resistant *Staphylococcus aureus* colonization and infection in a large teaching hospital in the Netherlands. *Infect Control Hosp Epidemiol,* 28**,** 970-5.

Vandenbergh, M. F., Yzerman, E. P., Van Belkum, A., Boelens, H. A., Sijmons, M. & Verbrugh, H. A. (1999) Follow-up of *Staphylococcus aureus* nasal carriage after 8 years: redefining the persistent carrier state. *J Clin Microbiol,* 37**,** 3133-40.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M. E. & Etienne, J. (2003) Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. *Emerg Infect Dis,* 9**,** 978-84.

Vaughn, C. P. & Elenitoba-Johnson, K. S. (2004) High-resolution melting analysis for detection of internal tandem duplications. *J Mol Diagn,* 6, 211-6.

Velappan, N., Snodgrass, J. L., Hakovirta, J. R., Marrone, B. L. & Burde, S. (2001) Rapid identification of pathogenic bacteria by single-enzyme amplified fragment length polymorphism analysis. *Diagn Microbiol Infect Dis,* 39, 77-83.

Verhoef, J., Beaujean, D., Blok, H., Baars, A., Meyler, A., Van Der Werken, C. & Weersink, A. (1999) A Dutch approach to methicillin-resistant *Staphylococcus aureus. Eur J Clin Microbiol Infect Dis,* 18, 461-6.

Vlack, S., Cox, L., Peleg, A. Y., Canuto, C., Stewart, C., Conlon, A., Stephens, A., Giffard, P., Huygens, F., Mollinger, A., Vohra, R. & Mccarthy, J. S. (2006) Carriage of methicillin-resistant *Staphylococcus aureus* in a Queensland Indigenous community. *Med J Aust,* 184, 556-9.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. & Et Al. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res,* 23, 4407-14.

Wagenlehner, F. M., Naber, K. G., Bambl, E., Raab, U., Wagenlehner, C., Kahlau, D., Holler, C., Witte, W., Weidner, W., Lehn, N., Harbarth, S. & Linde, H. J. (2007) Management of a large healthcare-associated outbreak of Panton-Valentine leucocidin-positive meticillin-resistant *Staphylococcus aureus* in Germany. *J Hosp Infect,* 67, 114-20.

Weese, J. S., Dick, H., Willey, B. M., Mcgeer, A., Kreiswirth, B. N., Innis, B. & Low, D. E. (2006) Suspected transmission of methicillin-resistant *Staphylococcus aureus* between domestic pets and humans in veterinary clinics and in the household. *Vet Microbiol,* 115, 148-55.

Weigel, L. M., Clewell, D. B., Gill, S. R., Clark, N. C., Mcdougal, L. K., Flannagan, S. E., Kolonay, J. F., Shetty, J., Killgore, G. E. & Tenover, F. C. (2003) Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus. Science,* 302, 1569-71.

Williams, R. E. (1963) Healthy carriage of *Staphylococcus aureus*: its prevalence and importance. *Bacteriol Rev,* 27, 56-71.

Williams, R. J., Ward, J. M., Henderson, B., Poole, S., O'hara, B. P., Wilson, M. & Nair, S. P. (2000) Identification of a novel gene cluster encoding staphylococcal exotoxin-like proteins: characterization of the prototypic gene and its protein product, SET1. *Infect Immun,* 68, 4407-15.

Wisplinghoff, H., Bischoff, T., Tallent, S. M., Seifert, H., Wenzel, R. P. & Edmond, M. B. (2004) Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin Infect Dis,* 39, 309-17.

Wisplinghoff, H., Rosato, A. E., Enright, M. C., Noto, M., Craig, W. & Archer, G. L. (2003) Related clones containing SCC*mec* type IV predominate among clinically significant *Staphylococcus epidermidis* isolates. *Antimicrob Agents Chemother,* 47, 3574-9.

Witte, W., Kresken, M., Braulke, C. & Cuny, C. (1997) Increasing incidence and widespread dissemination of methicillin-resistant *Staphylococcus aureus* (MRSA) in hospitals in central Europe, with special reference to German hospitals. *Clin Microbiol Infect,* 3, 414-422.

Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem,* 49, 853-60.

Woodford, N. (2001) Epidemiology of the genetic elements responsible for acquired glycopeptide resistance in *enterococci*. *Microb Drug Resist,* 7**,** 229-36.

Wu, C. Y., Hoskins, J., Blaszczak, L. C., Preston, D. A. & Skatrud, P. L. (1992) Construction of a water-soluble form of penicillin-binding protein 2a from a methicillin-resistant *Staphylococcus aureus* isolate. *Antimicrob Agents Chemother,* 36**,** 533-9.

Wu, S., Piscitelli, C., De Lencastre, H. & Tomasz, A. (1996) Tracking the evolutionary origin of the methicillin resistance gene: cloning and sequencing of a homologue of *mec*A from a methicillin susceptible strain of *Staphylococcus sciuri*. *Microb Drug Resist,* 2**,** 435-41.

Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., Yamada, S., Komatsuzawa, H. & Sugai, M. (2000) Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. *Mol Microbiol,* 38**,** 694-705.

Yang, J. A., Park, D. W., Sohn, J. W. & Kim, M. J. (2006) Novel PCR-restriction fragment length polymorphism analysis for rapid typing of staphylococcal cassette chromosome *mec* elements. *J Clin Microbiol,* 44**,** 236-8.

Yu, V. L., Goetz, A., Wagener, M., Smith, P. B., Rihs, J. D., Hanchett, J. & Zuravleff, J. J. (1986) *Staphylococcus aureus* nasal carriage and infection in patients on hemodialysis. Efficacy of antibiotic prophylaxis. *N Engl J Med,* 315**,** 91-6.

Zhang, K., Mcclure, J. A., Elsayed, S., Louie, T. & Conly, J. M. (2005) Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol,* 43**,** 5026-33.

Zulian, C., Descamps, P., Samyn, B., Lemerle, J. P. & Gaillot, O. (1999) [Inquiry into the incidence of nosocomial infections and evaluation of the transmission of methicillin-resistant *Staphylococcus aureus* in an orthopedic surgical unit]. *Pathol Biol (Paris),* 47**,** 445-8.